MIDDLEBURY COLLEGE



THE EGBERT STARR LIBRARY

David B. VanVleck

ARTHUR C. GUYTON, M.D.

PROFESSOR AND CHAIRMAN OF THE DEPARTMENT OF
PHYSIOLOGY AND BIOPHYSICS, UNIVERSITY OF MISSISSIPPI
SCHOOL OF MEDICINE

FUNCTION OF THE HUMAN BODY

SECOND EDITION

W. B. SAUNDERS COMPANY
Philadelphia and London

MIDDLEBURY COLLEGE
LIBRARY

3/1970
Gift

Science Center
QP
34
G8
1964
c.1

Reprinted October, 1964 and April, 1965

© 1964 by W. B. Saunders Company. Copyright 1959 by W. B. Saunders Company. Copyright under the International Copyright Union. All rights reserved. This book is protected by copyright. No part of it may be duplicated or reproduced in any manner without written permission from the publisher. Made in the United States of America. Press of W. B. Saunders Company. Library of Congress catalog card number 64-10657.

PREFACE

The immediate and continued acceptance of the first edition of this textbook, FUNCTION OF THE HUMAN BODY, has been very gratifying, a gratification that has driven me to devote much effort to the preparation of this second edition, to make it as accurate as possible and to improve the clarity wherever this could be done. In view of this goal, approximately one half of the book has been completely rewritten, and the remainder has been thoroughly revised. Forty-five new figures have been added, and over half of the remaining figures have been replaced or modified to improve their presentations.

Since the preparation of the first edition, tremendous progress has been made in all fields of physiology, but most particularly in the field of cell physiology. To keep abreast of the special developments in our knowledge of cellular function, several new chapters have been added, and emphasis has been given throughout the text to the close interrelationships between function of individual cells and overall function of the body.

The primary purpose of the text is to present the basic philosophy of the functioning systems in the human body, with the understanding that almost precisely these same principles also apply to function of any other mammal, and in many instances even to the function of still lower animals. The book is written especially for the student who requires a complete survey of mammalian physiology but does not need to know all the details proper to a reference text. An attempt has been made to present the subject in terms that will not require an extensive background in biology, physics, and chemistry; yet I have tried to maintain a completely scientific attitude throughout, hoping to leave the reader with as few vague generalizations as possible. For this reason, many very important and, in some instances, classic animal experiments are presented as illustrative examples of physiologic principles. Also, physiologists have learned much about bodily function from human beings themselves, for every abnormality that occurs in the body is in itself an experiment of nature which can often demonstrate as many physiologic principles as can an animal experiment in a research laboratory. Therefore, selected abnormalities are discussed as further examples of basic physiologic principles.

A tremendous amount of work by others in addition to myself has gone into preparing this text. I wish especially to thank many teachers who have used this text for their helpful suggestions in making improvements, which almost without exception have been worth while and have been accepted in preparing this revision. I wish, also, to thank the staff of W. B. Saunders Company for their continued help, Miss Grace Self and Mrs. Sue Saltvick for their excellent secretarial work, and Mrs. Carolyn Hull for drawing the majority of figures. We all hope that this revision will receive the same degree of acceptance as the first edition and that the many improvements will be of special benefit to the physiology courses for which the text was prepared.

ARTHUR C. GUYTON

Jackson, Mississippi

CONTENTS

SECTION ONE. INTRODUCTION: THE CELL

1. Introduction to Human Physiology 3
2. The Cell and Its Composition 12
3. Basic Functional Systems of the Cell 21
4. Control of Cell Function; and Cell Reproduction 32

SECTION TWO. THE BODY FLUIDS AND THE URINARY SYSTEM

5. The Fluids of the Body 43
6. Capillary Membrane Dynamics, and the Body's Special Fluid Systems 52
7. Formation of Urine by the Kidney, and Micturition 64
8. Regulation of Body Fluid Constituents and Volumes ... 75

SECTION THREE. BLOOD AND IMMUNITY

9. The Blood Cells 87
10. The Reticuloendothelial System, Immunity, and Allergy 100
11. Blood Coagulation, Transfusion, and Transplantation of Organs 107

SECTION FOUR. THE CARDIOVASCULAR SYSTEM

12. The Pumping Action of the Heart, and Its Regulation . . 119

13. Blood Flow through the Systemic Circulation and Its Regulation 131

14. Special Areas of the Circulatory System 141

15. Systemic Arterial Pressure and Hypertension 152

16. Cardiac Output, Venous Pressure, Cardiac Failure, and Shock 164

SECTION FIVE. RESPIRATION

17. Mechanics of Respiration and Transport of Oxygen and Carbon Dioxide 177

18. Regulation of Respiration and the Physiology of Respiratory Abnormalities 191

19. Aviation, Space, and Deep Sea Diving Physiology 200

SECTION SIX. THE NERVOUS SYSTEM AND MUSCLE

20. The Nerve, and Membrane Potentials 215

21. Muscle Physiology 225

22. Synaptic Functions of Neurons, and Design of the Nervous System 236

23. Somesthetic Sensations and Interpretation of Sensations by the Brain 246

24. The Thought Processes and Control of Motor Activities 257

25. Reflex Functions of the Spinal Cord and Brain Stem . . . 270

26. The Autonomic Nervous System, Sleep, and Psychosomatic Effects 280

SECTION SEVEN. THE SPECIAL SENSORY SYSTEMS

27. The Eye 297

28. Hearing, Taste, and Smell 309

SECTION EIGHT. THE GASTROINTESTINAL AND METABOLIC SYSTEMS

29. Gastrointestinal Movements and Secretion, and Their Regulation 321

30. Digestion and Assimilation of Carbohydrates, Fats, and Proteins 334

31. Release of Energy from Foods, and Nutrition 347

32. Body Heat, and Temperature Regulation 358

SECTION NINE. ENDOCRINOLOGY AND REPRODUCTION

33. Introduction to Endocrinology: The Hypophyseal Hormones and Thyroxine 371

34. Adrenocortical Hormones and Insulin 379

35. Calcium Metabolism, Bone, Parathyroid Hormone, and Physiology of Teeth 387

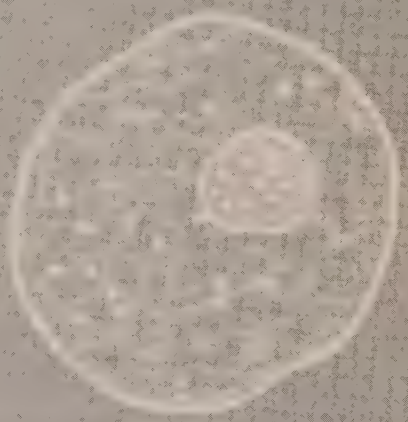
36. Sexual Functions of the Male and Female, and the Sex Hormones 397

37. Reproduction and Fetal Physiology 407

INDEX 419

Section One

INTRODUCTION: THE CELL



INTRODUCTION TO HUMAN PHYSIOLOGY

1

Physiology is the study of function in living organisms. It attempts to discover answers to such questions as: How and why do plants grow? What makes bacteria divide again and again? How do fish obtain oxygen from the sea and what use do they make of it? How is food digested? What is the nature of the thinking process in the brain? How do reproductive systems function?

Even small viruses weighing a million times less than bacteria have the characteristics of life, for they feed on their surroundings, they grow and reproduce, and they excrete by-products. These very minute living structures are the subject of the simplest type of physiology, *viral physiology.* Physiology becomes progressively more complicated and vast as it extends through the study of higher and higher forms of life such as bacteria, cells, plants, lower animals, and, finally, human beings. It is obvious, then, that the subject of this book, "human physiology," is but a small part of this vast discipline.

As small children we begin to wonder what enables people to move, how it is possible for them to talk, how they can see the expanse of the world and feel the objects about them, what happens to the food they eat, how they derive from food the energy for locomotion and other types of bodily activity, by what process they reproduce other beings like themselves so that life goes on, generation after generation. All these and other human activities make up *life.* Physiology attempts to explain them and hence to explain life itself.

ORGANIZATION OF THE HUMAN BODY

The human body has two basic levels of organization: the cellular level and the organ level. The cells are the building blocks of the organs, and the organs perform specialized functions for the entire body. Many human physiologists like to think that the cell is of minor importance in the total function of the human being. However, one will begin to appreciate the importance of the cell when he realizes that equally as many millions of years went into the evolutionary development of the cell as into its further development into the multi-trillion cellular aggregate that we call the human being. Therefore, an understanding of the cell and of cellular function is a prerequisite to understanding how the organs work and how these in turn function together to maintain the life of the human body.

Each cell is a living unit in itself, capable of existing, performing chemical reactions, contributing its part in the overall function of the body, and reproducing itself when its number in the body is too low. The only requirement for it to continue all these functions is that the constituents of the fluids in the spaces between the cells be controlled very exactly. Indeed, cells can even live after removal from the body if they are placed in a fluid bath that contains the same constituents and has the same physical conditions as those found in the body fluids.

The functions of all the organs in the body are directed toward a single goal, to maintain constant conditions in the fluids that surround the cells, thereby allowing the cells to continue living, growing, and reproducing. Claude Bernard, the great French physiologist who originated much of our modern physiological thought, called these fluids the *milieu interne,* which means the "internal environment," and Walter Cannon, another great physiologist of the first half of this century, referred to the maintenance of constant conditions in these fluids as *homeostasis.*

Every organ, with the exception of the reproductive organs, is concerned either directly or indirectly with homeostasis. For instance, the *circulatory system,* composed of the *heart* and *blood vessels,* transports blood throughout the body; water and dissolved substances diffuse back and forth between the blood and the fluids that surround the cells. Thus, the circulatory system keeps the fluids of all parts of the body constantly mixed. This function of the circulatory system is so effective that hardly any portion of fluid in any part of the body remains unmixed with the other fluids of the body more than a few minutes at a time. The *respiratory system* transfers oxygen from the air to the blood, and the blood in turn transports the oxygen to all the tissue spaces surrounding the cells. Carbon dioxide excreted by the cells enters the tissue fluids, then becomes mixed with the blood and is finally removed through the lungs. The *digestive system* performs a similar function for other nutrients besides oxygen; it processes nutrients which are then absorbed into the blood and are rapidly transported throughout the body fluids. The *liver,* the *endocrine glands,* and some of the other organs participate in what is collectively known as *intermediary metabolism,* which converts many of the nutrients absorbed from the gastrointestinal tract into substances that can be used directly by the cells. The *kidneys* remove the remains of the nutrients after their energy has been extracted by the cells, and other organs provide for *hearing, feeling, tasting, smelling,* and *seeing,* all of which aid the animal in his search for and selection of food and also help him to protect himself from dangers so that he can perpetuate the almost Utopian internal environment in which his cells continue their life processes.

Thus, all the organ functions of the body depend on individual functions of the cells, and, in turn, sustained life of the cells depends on maintenance of an appropriate fluid environment.

Cellular function is so important to the understanding of all phases of human physiology that a detailed consideration of this subject begins in the following chapter. The remainder of this chapter reviews the basic organ systems of the human body so that those beginning students of physiology who might not have studied animal anatomy can begin to see how the cells and their supporting structures are aggregated into a total functioning system.

The Skeleton and Its Muscles

Figure 1 illustrates the skeleton with some of its muscles still attached. Each joint of the skeleton is enveloped by a loose *capsule,* and the space within the capsule and between the two respective bones is the *joint cavity.* In the joint cavities is a thick, slippery fluid containing hyaluronic acid, which lubricates the joints, promoting ease of movement. On the sides of each capsule are strong fibrous *ligaments* that keep the joints from pulling apart. Often the ligaments are only on two sides of the joint, which allows the joint to move freely in one direction but not so freely in another direction. Other joints, particularly those of the

spine, hips, and shoulders, not having very restrictive ligaments, can move in almost any direction; that is, they can bend forward, backward, and to either side, or they can even be rotated. In these instances, loose ligaments merely limit the degree of motion to prevent excessive movement in any one direction.

Muscles move the limbs and other parts of the body in all directions allowed by the ligaments. In the case of movement at the knee joint, for instance, one major muscle is on the front and several muscles are on the back of the joint. Contraction of the anterior muscle pulls the lower leg forward, while contraction of the posterior muscles pulls it backward. There is a similar arrangement of muscles anteriorly and posteriorly about the ankle, except that the ligaments of the ankle allow the ankle joints to move also from side to side, and additional muscles are available to provide the sidewise movements. The muscles of the spine are especially interesting because, contrary to what might be expected, the back muscles are not just a few very large muscles, but are composed of about one hundred different individual muscles each one of which performs a specific function: one rotates an adjacent vertebra while a second one flexes the vertebra forward, a third backward, and so on. This is analogous to the arrangement of the centipede, for each segment can bend independently of all the others. The joint connecting the head to the spinal column is also supplied with many separate muscles arranged on all sides so that the head can be rotated from side to side or bent in any direction.

In summary, then, the skeleton is literally a bag of bones which can be contorted into many different configurations. Each bone has its own function, and the limitations of angulation of each joint are decreed by the ligaments. The knee joint bends in only one direction, the ankle joint in two, the hip joint in two directions plus an additional rotary motion; and, in general, at least two muscles are available for each motion that the ligaments of a joint allow, one for each of the two directions of movement.

The muscles themselves are composed of long *muscle fibers.* Usually many thousands of these fibers are oriented side by side like the threads in a skein of wool. At each end of the muscle, the muscle fibers fuse with strong *tendon fibers* that form a bundle called the *muscle tendon.* The muscle tendons in turn penetrate and fuse with the bones on the two sides of the respective joint so that any pull exerted will effect appropriate movement.

All muscles are not exactly alike in size and appearance; for instance, the smallest skeletal

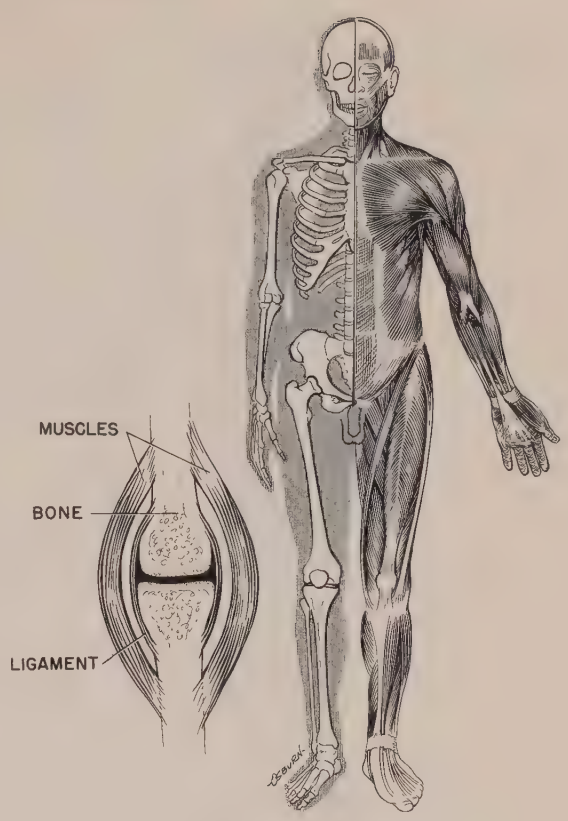


Figure 1. The skeletal and muscular systems of the human body.

muscle of the body, the stapedius, a minute muscle in the inner ear, is only a few millimeters long, while the longest muscle, the sartorius, extends almost two feet down the entire length of the thigh, connecting the pelvic bone with the lower leg. Some muscles such as those of the abdominal wall, are arranged in thin sheets, while others are round, cigar-shaped structures, for example, the biceps which lifts the lower arm and the gastrocnemius which pulls the foot down when one wishes to stand on his tiptoes.

The precise method by which muscle fibers contract is still not completely clear, but we do know that signals arriving in the muscles through nerves cause each fiber to shorten for a brief instant, allowing the entire muscle belly to contract and thereby to perform its function.

The Nervous System

The nervous system, illustrated in Figure 2, is composed of the brain, the spinal cord, and the peripheral nerves that extend throughout the body. A major function of the nervous system is

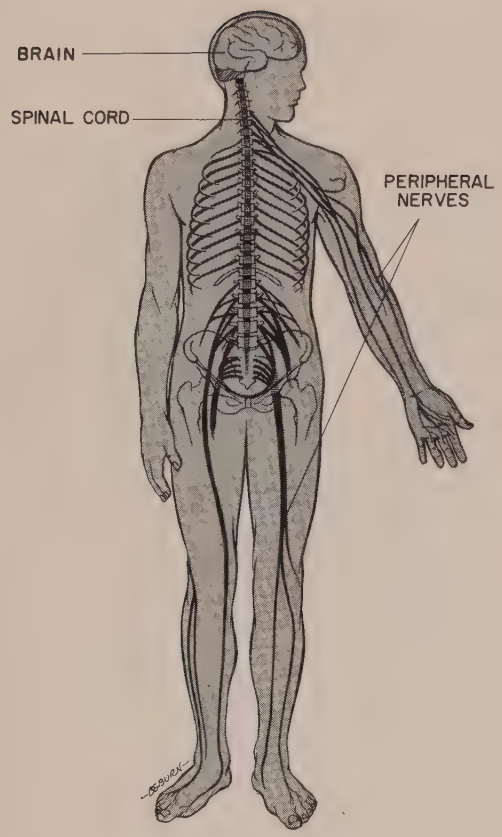


Figure 2. The nervous system.

to control many of the bodily activities, especially those of the muscles, but to exert this control intelligently the brain must be apprised continually of the body's surroundings. To perform these varied activities, the nervous system is composed of two separate portions, the *sensory portion* which reports and analyzes the nature of conditions around the body, and the *motor portion* which controls the muscles.

The sensory portion operates through the senses of sight, hearing, smell, taste, and feel. The sense of feel is actually many different senses, for one can feel light touch, pin pricks, pressure, pain, vibration, position of the joints, tightness of the muscles, and tension on the tendons.

Once information has been relayed to the brain from all the senses, the brain then determines what movement, if any, is most suitable, and the muscles are called into action to implement the decision.

One of the most important functions of the nervous system is to regulate walking. In walking, the body must be supported against gravity, the legs must move rhythmically in a walking motion, equilibrium must be maintained, and the direction of movement of the limbs must be guided. Therefore, the initiation and control of locomotion are very complex functions of the nervous system and require the services of a major portion of the brain.

The autonomic nervous system. The autonomic nervous system, which is really part of the motor portion of the nervous system, regulates many of the internal functions of the body. It operates principally by causing contraction or relaxation of a type of muscle called *smooth muscle*, which constitutes the major portion of many of the internal organs. Smooth muscle fibers are much smaller than skeletal muscle fibers, and they usually are arranged in large muscular sheets. For instance, the gastrointestinal tract, the urinary bladder, the uterus, the biliary ducts, and the blood vessels are all composed mainly of smooth muscle sheets rolled into tubular or spheroid structures. Some of the autonomic nerves cause the muscles of these organs to contract while others cause relaxation.

The autonomic nerves also regulate secretion by many of the glands in the gastrointestinal tract and elsewhere in the body, and at times their nerve endings even secrete hormones that can increase or decrease the rates of chemical reactions in the body's tissues.

Finally, the autonomic nervous system helps to control the heart, which is composed of *cardiac muscle*, a type of muscle intermediate between smooth muscle and skeletal muscle. Stimulation of the so-called *sympathetic fibers* causes the rate and force of contraction of the heart to increase, whereas stimulation of the *parasympathetic* fibers causes the opposite effects.

In summary, the autonomic nervous system helps to control most of the body's internal functions.

The Circulatory System

The circulatory system, illustrated in Figure 3, is composed mainly of the heart and blood vessels. The heart consists of two separate pumps arranged side by side. The first pump forces blood into the lungs. From them, the blood returns to the second pump to be forced then into the *systemic arteries* which transport it through the body. From the arteries it flows into the *capillaries*, then into the *veins* and finally back to the heart, thus making a complete circuit. Circulating around and around through the body, the blood acts as a transportation system for conducting various chemical substances from one place to another. It is the circulatory system that carries

nutrients to the tissues and then carries excretory products away from the tissues.

The capillaries are porous, allowing fluid and nutrients to diffuse into the tissues and excreta from the cells to reenter the blood.

Large particles, such as old debris of dead tissues, protein molecules, and dead bacteria cannot pass from the tissues directly into the blood through the small pores of the capillaries. A special accessory circulatory system known as the *lymphatic system* takes care of these materials. Lymphatic vessels originate in small *lymphatic capillaries* which lie beside the blood capillaries, and *lymph*, which is fluid derived from the spaces between the cells, flows along the lymphatic vessels up to the neck where these vessels empty into the neck veins. The lymphatic capillaries are extremely porous so that large particles can enter the lymphatic system and be transported by the lymph. At several points the lymphatic vessels pass through *lymph nodes* where most large particles are filtered out and where bacteria are engulfed and digested by special cells called *reticuloendothelial* cells.

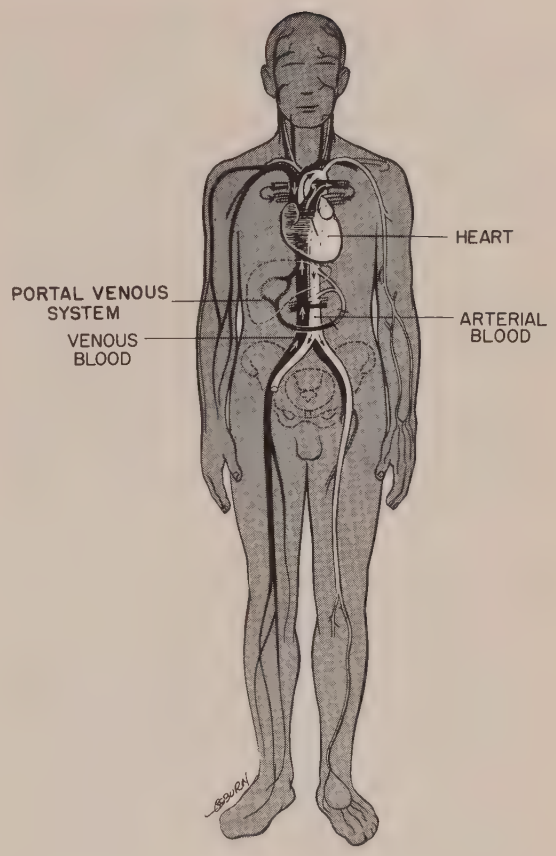


Figure 3. The circulatory system: heart and major vessels.

The Respiratory System

Figure 4 is a diagram of the respiratory system, showing the two fundamental portions of this system: (1) the air passages and (2) the blood vessels of the lungs. Air is moved in and out of the lungs by contraction and relaxation of the respiratory muscles, and blood flows continually through the vessels. Only a very thin membrane separates the air from the blood, and since this membrane

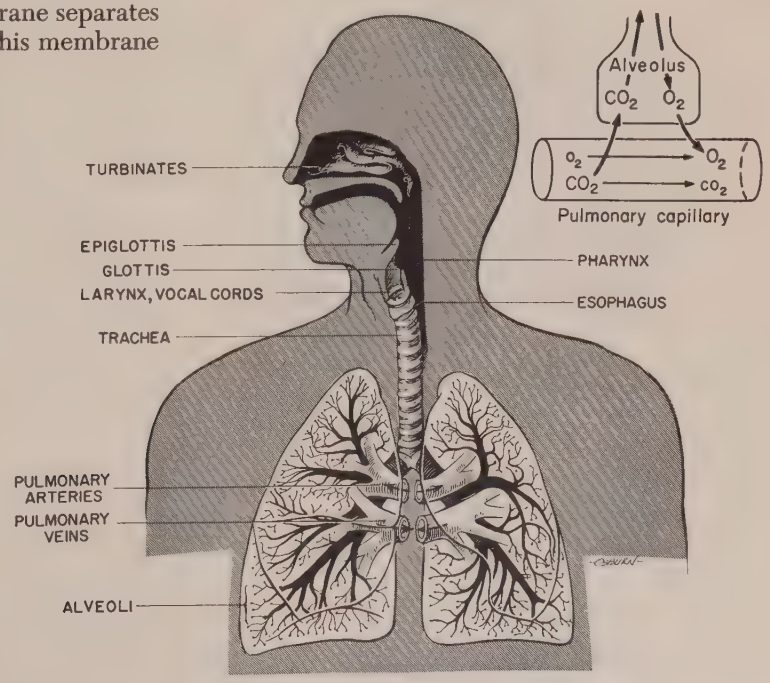


Figure 4. The respiratory system.

is porous to gases it allows free passage of oxygen into the blood and of carbon dioxide from the blood into the air.

Oxygen is one of the nutrients needed by the body's tissues. It is carried by the blood and tissue fluids to the cells where it combines chemically with foods to release energy. This in turn is used to promote muscle contraction, secretion of digestive juices, conduction of signals along nerve fibers, and synthesis of many substances needed for growth and function of cells.

When oxygen combines with foods to liberate energy, carbon dioxide is formed. This is carried by tissue fluids to the blood and by the blood to the lungs. Then, the carbon dioxide diffuses from the blood into the lung air to be breathed out into the atmosphere.

The Gastrointestinal System

The digestive system is illustrated in Figure 5. Food after being swallowed enters the stomach, then the duodenum, the jejunum, the ileum, and the large intestine, finally to be excreted through the anus. Those portions of the food valuable to the body are chemically and physically extracted by the process of *digestion,* and are transferred into the blood by the process of *absorption.* Along the entire extent of the gastrointestinal tract, special substances are secreted into the gut either continuously or when food is present. These secretions contain *digestive enzymes* which cause the foods to split into chemicals small enough to pass through the pores of the intestinal membrane into underlying capillary and lymphatic vessels. Thence the digestive products enter the circulating blood to be transported and used where they may be needed.

Metabolic Systems

Metabolism and growth. The term *metabolism* means simply the chemical reactions that occur in the animal organism. These reactions occur inside the individual cells which make up the tissues, and their functions are to provide energy to perform the bodily activities and to build new structures. It is because of the metabolic processes that the cells grow larger and more numerous. The metabolism of special cells allows them to form structures such as bones and fibrous tissue, enlarging the entire animal. Thus, metabolism is the basis not only for the energy needed by the body but for growth itself.

INTERMEDIARY METABOLISM. Many of the foods entering the blood from the digestive tract

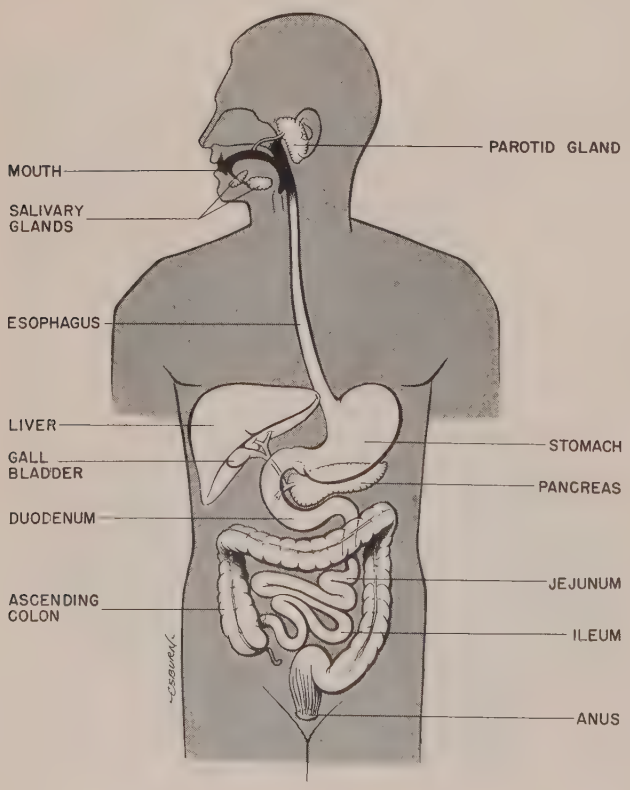


Figure 5. The digestive system.

can be used by the body's tissues without alteration, but some tissues require special chemicals which are not normally found in the food. To supply these, much of the absorbed food passes to special organs where it is changed into new substances needed by the cells. This process is called *intermediary metabolism.*

The liver. The liver is one of the internal organs especially adapted for intermediary metabolism and storage. It can split fats and proteins into smaller substances so that the tissues can use them for energy, and it forms products needed for blood coagulation, for transport of fat, for immunity to infection, and for many other purposes. The liver is also capable of storing large quantities of fats, carbohydrates, and even proteins and then releasing these foods when the tissues need them. An animal can live for only a few hours without a liver.

Control of metabolism by the hormones. Metabolism is an inherent function of every cell of the body. However, the rate of metabolism in each respective cell is very often increased or decreased by the controlling action of *hormones* secreted by *endocrine glands* in different parts of the animal. The *thyroid gland,* located in the neck, secretes *thyroxine* which acts on all cells of the body to increase the rate of most metabolic reactions. *Epinephrine* and *norepinephrine,* two hormones secreted by the *adrenal medullas,* also increase the rate of metabolism in all cells. The *ovaries* secrete *estrogens* and *progesterone,* and the *testes* secrete *testosterone,* which help to control metabolism in the sex organs of the female and male, respectively. *Insulin,* secreted by the *pancreas,* a gland located behind and beneath the stomach, increases the utilization of carbohydrates and decreases the utilization of fats in all the tissues. *Adrenocortical hormones* secreted by the two *adrenal cortices,* located at the upper poles of the kidneys, help to convert proteins to carbohydrates, and they control the passage of proteins, salts, and perhaps other substances through the cell walls. Finally, *parathyroid hormone,* secreted by four minute *parathyroid glands* located behind the thyroid gland in the neck, controls the amount of calcium in the blood by removing calcium from the bones when it is needed and by allowing more deposition in the bones when it is not needed.

The Excretory System

The *kidneys,* illustrated in Figure 6, constitute an excretory system for ridding the blood of unwanted substances. Most of the substances are the end-products of metabolic reactions, and they include mainly urea, uric acid, creatinine, phenols, sulfates, and phosphates. If they were allowed to collect in the blood in large quantities, the "ashes of the cellular fires" would soon "smother the flames" themselves so that no further metabolic reactions could take place. For this reason it is important that the kidneys remove these unwanted substances.

The kidneys have another very valuable function besides that of excretion: they regulate the concentrations of most of the electrolytes in the body fluids. A very large proportion of these electrolytes is sodium chloride, common table salt. The kidneys continually adjust the concentrations of both sodium and chloride, and they also regulate very precisely the concentrations of potassium, magnesium, phosphates, and many other substances. The kidneys perform this function by allowing the unwanted substances such as urea to pass easily into the urine while retaining the wanted substances such as glucose. If sodium is already present in the blood in too large a concentration, it becomes an unwanted substance, and much of it is excreted by the kidneys into the urine, but, if the concentration of sodium is too low, it is a wanted substance instead, and the kidney's special properties, which will be described in a future chapter, then prevent its loss from the blood.

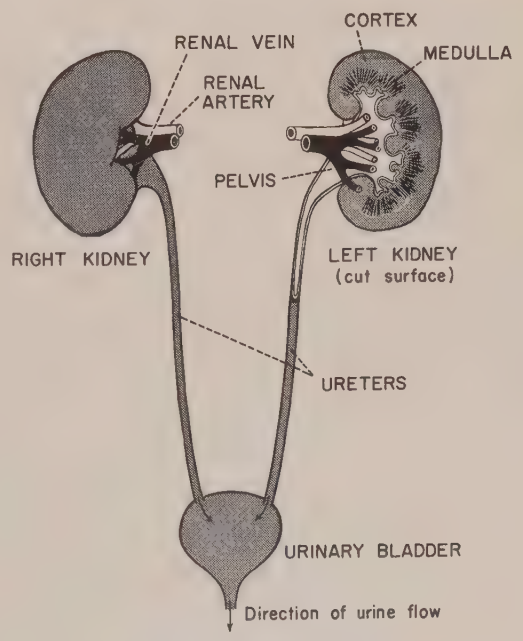


Figure 6. The kidneys and urinary system.

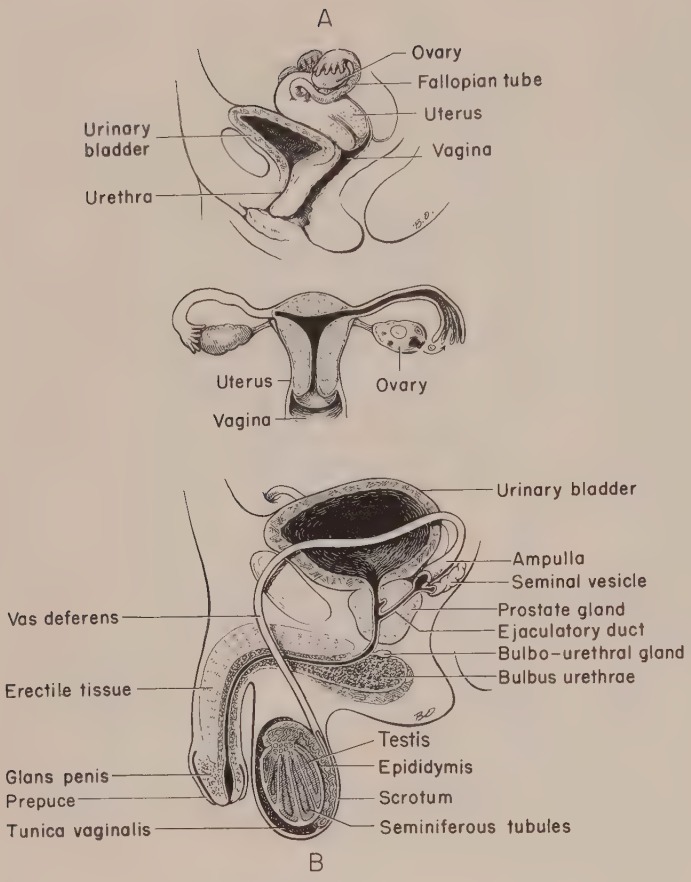


Figure 7. The reproductive systems: (A) female, (B) male.

The Reproductive Systems

All the functions and systems of the body that maintain life would be useless were it not for those that provide for its reproduction. The reproductive systems, shown in Figure 7, are different in the female and male. The female (Fig. 7A) provides the *ovum* (egg) from which a new human being is to develop, but this ovum cannot begin developing until it is fertilized by a *sperm* from the male (Fig. 7B). The fertilized ovum derives half of its developmental characteristics from the mother and half from the fertilizing sperm of the father, so that the offspring owes its characteristics equally to each of the parents.

After the ovum has been fertilized it is at first still a single cell, but soon it divides into two cells, four cells, and finally into many cells, thus becoming an *embryo*. Gradually, the newly developing cells *differentiate* into the organs of the body.

The mother provides nutrition for the growing embryo by means of the *placenta*, a structure attached to the inner wall of her *uterus*. Nutrients diffuse from the mother's blood into the baby's blood through the *placental membrane* which is very much like the membrane of the respiratory system. In turn, excretory products from the baby pass into the mother's blood. Thus, the embryo is nurtured through a period of nine months in the mother's body until it becomes capable of sustaining life on its own in the outer world. At that time the mother's uterus expels the baby.

COMMUNAL ORGANIZATION OF THE BODY

If this chapter has succeeded in presenting a résumé of the functioning systems of the human body, it should by now be obvious that no single part of the human body can live by itself, but that each of the body's functions is necessary for continuous operation of the others. The human animal is a sensing, thinking, and motile organism which, by virtue of nervous and hormonal systems of control, can adapt itself to most surroundings. Its activities are initiated and controlled in part

involuntarily, in part by intuition, and in part by reasoning. In the framework of the organs and other tissues are literally billions of individual *cells* each one of which is a living structure. It is the magic of these cells which makes the human body possible. The next chapter describes the function of the cell itself.

GENERAL REFERENCES IN THE FIELD OF PHYSIOLOGY

American Journal of Physiology. This journal, published monthly, contains articles by authors throughout the world on current research projects.

Annual Review of Physiology, Palo Alto, Calif., Annual Reviews, Inc. (One book each year.) Each book contains review articles that cover the literature of the preceding year in almost the entire field of animal physiology.

Bard, P.: *Medical Physiology.* 11th Ed. St. Louis, The C. V. Mosby Company, 1961. This text is most useful for postgraduate students.

Best, C. H., and Taylor, N. B.: *The Physiological Basis of Medical Practice.* 7th Ed. Baltimore, The Williams & Wilkins Company, 1961. This text is written principally for the postgraduate student.

Guyton, A. C.: *Textbook of Medical Physiology.* 2nd Ed. Philadelphia, W. B. Saunders Company, 1961. This text presents physiology at the level of the medical student. In general it covers the same material as that in the present text, but in much greater detail and with more emphasis on the medical aspects of human physiology.

Physiological Reviews. This is a journal published four times a year that contains reviews of most subjects in the field of physiology every few years.

Ruch, T. C., and Fulton, J. F.: *Medical Physiology and Biophysics.* 18th Ed. Philadelphia, W. B. Saunders Company, 1960. This text is written at the level of the postgraduate student.

THE CELL AND ITS COMPOSITION

2

The human body contains about 100 trillion cells, each of which is a living structure. Since the functions of the body's organs are performed by their constituent cells, the overall function of the human body is actually the combined function of all these 100 trillion cells.

Several hundred types of cells exist in the human body, and each type plays a special role in bodily function. Yet, despite this difference between cells, they all have some functions in common, for instance, their abilities to live, grow, and reproduce. The purpose of this chapter is mainly to emphasize these similarities of cells and their functions. Yet, first, let us describe some of the different types of cells and their related structures.

Some Representative Types of Cells

Figure 8 illustrates five types of tissues that perform different functions in the body. Example A depicts *connective tissue* which holds the different structures of the body together. This tissue contains cells called *fibroblasts* that are enmeshed in *collagenous* and *elastic* fibers. The fibroblasts secrete substances that polymerize to form the fibers, and the fibers in turn provide tensile strength to the tissues, thereby holding them together.

Example B illustrates several *red* and *white blood cells.* The red cells carry oxygen from the

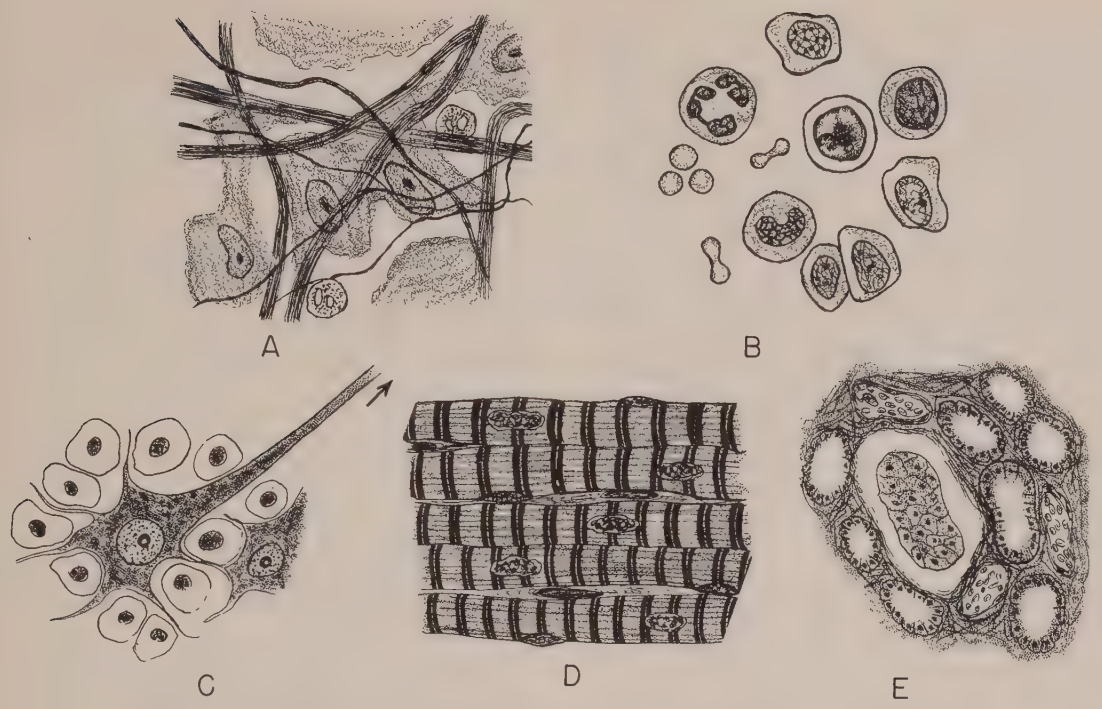


Figure 8. Examples of different types of cells: (A) connective tissue, (B) red and white blood cells, (C) neuronal tissue, (D) muscle cells, (E) kidney tissue.

lungs to the tissues and carbon dioxide from the tissues back to the lungs, while the white blood cells cleanse the blood and tissues of unwanted materials such as bacteria, debris from degenerating tissues, and so forth.

Example C shows a *nerve cell* and surrounding *supporting cells* in the brain substance. The long projection of the nerve cell is its *axon* that often extends as long as one meter. Electrochemical impulses travel over the surface of the nerve cell and along the membrane of the axon, thereby transmitting information from one part of the body to another.

Example D illustrates *muscle cells* which can also transmit electrochemical impulses over their membranes but which are different from nerve cells in that they contain long myofibrils that extend the entire length of the muscle and contract when an electrochemical impulse travels over the surface of the fibril.

Example E illustrates several different types of cells and structures in the *kidney*. Connective tissue is present throughout the kidney to hold the different structures together. Several *kidney tubules* are shown, lined by *epithelial cells;* these structures help to form the urine, as is explained in a later chapter. Also, two small *blood vessels* are illustrated in cross-section in this figure; these vessels are filled with red blood cells.

The representative tissues and cells shown in Figure 8 are but a few of the many types found in the body, but they show the wide variability between different cells. These dissimilarities allow cells to perform different functions. The remainder of this chapter, however, presents the *similarities* between cells rather than their dissimilarities. In future chapters, many different types of specialized cells are described in detail, and their functions are presented.

Comparison of the Animal Cell With Precellular Forms of Life

Many of us think of the cell as the lowest level of animal life. However, the cell is a very complicated organism, which probably required several billion years to develop after the earliest form of life, the *virus,* first appeared on Earth. Figure 9 illustrates the relative sizes of the *smallest known virus,* a *large virus,* a *rickettsia,* a *bacterium,* and the *cell,* showing that the cell has a diameter about 1000 times that of the smallest virus, and, therefore, a volume about 1 billion times that of the smallest virus. Correspondingly, the functions and anatomical organization of the cell are far more complex than those of the virus.

The principal constituent of the very small virus

is a substance called *nucleic acid.* This is an *autocatalytic substance,* which means that it is capable of reproducing itself if appropriate nutrients are available. Thus, the virus is capable of propagating its lineage from generation to generation, and, therefore, is a living structure in the same way that the cell and the human being are living structures.

As life evolved to the large viruses, other chemicals besides nucleic acid became a part of the organism, and specialized functions began to develop in different parts of the virus. A *membrane* formed around the virus, and inside the membrane a fluid *matrix* appeared. Specialized chemicals developed inside the matrix to perform special functions; *protein enzymes* appeared which were capable of catalyzing chemical reactions and, therefore, of controlling the organism's activities.

In the rickettsial and bacterial stages, *organelles* developed, representing aggregates of chemical compounds that perform functions in a more efficient manner than can be achieved by dispersed chemicals throughout the fluid matrix. And, finally, in the cell stage more complex organelles developed, the most important of which is the *nucleus.* The nucleus distinguishes the cell from all lower forms of life; this structure provides a control center for all cellular activities, and it also provides for very exact reproduction of new cells generation after generation, each new cell having essentially the same structure as its progenitor.

PROTOPLASM AND ITS CHEMICAL NATURE

The term *protoplasm* means simply the substance of the cell. Thus, protoplasm includes both *cytoplasm* and *nucleoplasm,* the constituents of the cytoplasmic and nuclear compartments of the cell. In some usages of the word protoplasm, even the cell and nuclear membranes are included.

Protoplasm is composed mainly of five major basic substances: (1) water, (2) electrolytes, (3) proteins, (4) lipids, and (5) carbohydrates.

Water. The fluid medium of all protoplasm is water, which is present in a concentration between 70 and 85 per cent. Many cellular chemicals are dissolved in the water, while others are suspended in small particulate form. Chemical reactions take place between the dissolved chemicals or at the surface boundaries of the suspended particles. The fluid nature of the water allows

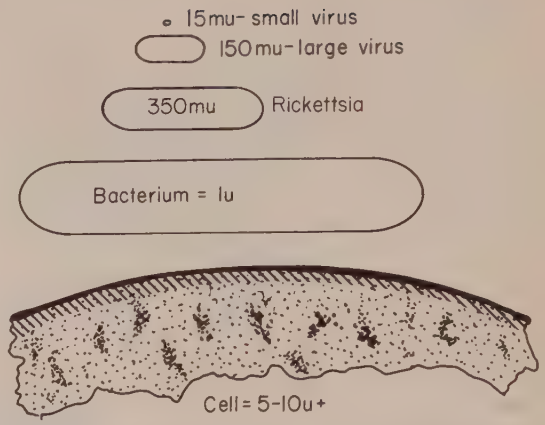


Figure 9. Comparison of sizes of subcellular organisms with that of the average cell in the human body.

both the dissolved and suspended substances to diffuse to different parts of the cell, thereby providing transport of the substances from one part of the cell to another.

Electrolytes. The most important electrolytes in the cell are *potassium, magnesium, phosphate, sulfate, bicarbonate,* and small quantities of *sodium* and *chloride.* These will be discussed in much greater detail in Chapter 5 which will consider the interrelationships between the different fluids of the body.

The electrolytes are dissolved in the water of protoplasm, and they provide inorganic chemicals for cellular reactions. Also, they are necessary for operation of some of the cellular control mechanisms. For instance, electrolytes acting at the cell membrane allow transmission of electrochemical impulses in nerve and muscle fibers, and the intracellular electrolytes determine the activity of different enzymatically catalyzed reactions that are necessary for cellular metabolism.

Proteins. Next to water, the most abundant substance in most cells is the proteins, which normally constitute 10 to 20 per cent of the cell mass. These can be divided into two different types, *structural* and *enzyme proteins.*

To give one an idea of what is meant by *structural proteins,* one needs only to note that hair is almost entirely a structural protein. Proteins of this same general type are present in the cell membranes, in the nuclear membranes, and in membranes surrounding special intracellular structures such as the endoplasmic reticulum, the mitochondria, and so forth. Thus, the structural proteins hold the structures of the cells together. Most structural proteins are *fibrillar;* that is, the individual protein molecules are polymerized into long fibrous threads. The threads in turn provide tensile strength for the cellular structures.

The *enzymes,* on the other hand, are an entirely different type of protein, composed usually of individual protein molecules or at most of aggregates of a few molecules in a globular form rather than fibrillar form. These proteins, in contrast to the fibrillar proteins, are often soluble in the fluid of the cell or are adsorbed to the surface of membraneous structures inside the cell. The enzymes come into direct contact with other substances inside the cell and catalyze chemical reactions. For instance, the chemical reactions that split glucose into its component parts and then combine these with oxygen to form carbon dioxide and water are catalyzed by a series of protein enzymes. It is in this way that enzyme proteins control the metabolic functions of the cell.

Special types of enzyme proteins are present in different parts of the cell. Of particular importance are the *nucleoproteins,* present both in the nucleus and in the cytoplasm. The nucleoproteins of the nucleus constitute mainly the *genes* that control both the overall function of the cell and also control heredity from cell to cell. These substances are so important that they will be considered in detail in Chapter 4.

The chemical nature of proteins will be considered in detail in Chapter 30, and the different structural and enzymatic functions of proteins will be subjects of discussion at numerous points throughout this text.

Lipids. Lipids are several different types of substances that are grouped together because of their common property of being soluble in fat solvents. For instance, the usual fat of animal tissues is a lipid called *neutral fat.* However, in addition to neutral fat two other types of lipids, *phospholipids* and *cholesterol,* are very common throughout the cells.

The usual cell contains 2 to 3 per cent lipids which are dispersed throughout the cell but are present in especially high concentrations in the cell membrane, the nuclear membrane, and the membranes lining intracytoplasmic organelles.

The special importance of lipids in the cell is that they are either insoluble or only partially soluble in water. They combine with structural proteins to form the different membranes that separate the different water compartments of the cell from each other. These membranes contain approximately 40 per cent lipids combined with structural proteins. The lipids of each membrane form a boundary between the solutions on the two sides of the membrane, making it impervious to many dissolved substances.

Some lipids are partially soluble in water as well as in lipid solvents. For instance, the phospholipid molecule has a *hydrocarbon portion* that makes it soluble in lipid solvents and a *positively charged polar portion* that makes it soluble in water. Therefore, phospholipid molecules aggregate at the surfaces of the different membranes, with part of their structure dissolved in the fat matrix of the membrane and the other part dissolved in the water.

In summary, the special characteristics of lipids allow the cell to develop many membranous structures that can either limit the transfer of water soluble substances from one part of the cell to another or can selectively pass substances if they meet appropriate chemical criteria as will be discussed in Chapter 3. The chemical natures of the different types of fats and their functions in the body will be discussed in Chapter 30.

Carbohydrates. In general, carbohydrates have very little structural function in the cell, but they play a major role in cellular nutrition. Most human cells do not maintain large stores of carbohydrates, averaging about 1 per cent of their total mass. However, carbohydrate, in the form of *glucose,* is always present in the surrounding extracellular fluid so that it is readily available to the cell. The small amount of carbohydrate stored in the cells is almost entirely in the form of *glycogen,* which is an insoluble polymer of glucose.

Glucose is one of the major nutrients required for cellular function. Enzymes present throughout the cytoplasm and even in the nucleus split glucose into *smaller hydrocarbons,* thereby releasing small amounts of energy that can be used to energize cellular functions. The smaller hydrocarbons then enter the *mitochondria* where they react with oxygen to form *carbon dioxide* and *water.* In this second process, tremendous quantities of energy are released. This energy in turn is utilized by the cell to perform the different cellular functions such as synthesis of new chemical compounds, muscular contraction, transmission of nerve impulses, and so forth.

PHYSICAL STRUCTURE OF THE CELL

The cell is not merely a bag of fluid, enzymes, and chemicals, but has highly organized physical structures called *organelles,* which are equally as important to the function of the cell as the cell's chemical constituents. For instance, if one group of organelles, the *mitochrondria,* should be absent, the energy-requiring functions of the cell would cease entirely. In the following pages the most important of the cellular organelles are described

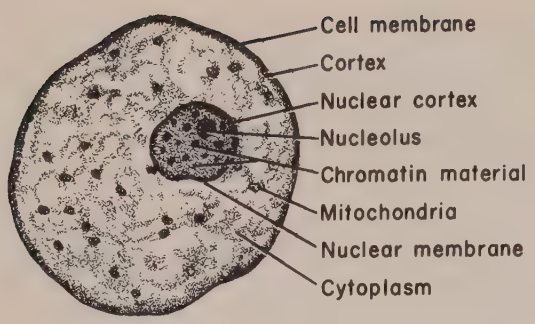


Figure 10. Structure of the cell as seen by the light microscope.

and their roles in cellular function, are briefly mentioned.

General Physical Organization of the Cell

Figure 10 illustrates the general organization of the cell as it is seen under the light microscope. One readily notes that the cell is divided into two major compartments, the *cytoplasm* and the *nucleus*, and that these are bounded on their outsides by the *cell membrane* and *nuclear membrane* respectively. Also, the light microscope shows a number of *inclusions* in both the cytoplasm and the nucleus. However, the fine structures of these inclusions cannot be discerned with this type of microscope. Only within the last 20 to 30 years, since the advent of the electron microscope, have we been able to determine the details of cellular ultrastructure and to make an attempt at understanding the basic mechanisms of intracellular function. Let us now present some of these details.

The Cell Membrane

The cell membrane is a thin elastic structure approximately 75 to 100 angstroms in thickness. It is composed almost entirely of proteins and lipids, with an approximate percentage composition of 60 per cent protein and 40 per cent lipids. The proteins in the cell are mainly a type of protein called *stromatin*, an insoluble structural protein having elastic properties. The lipids are approximately 65 per cent phospholipids, 25 per cent cholesterol, and 10 per cent other lipids.

The precise molecular organization of the cell membrane is not yet clear, but many experiments point to the structure illustrated in Figure 11A. This shows a central layer of lipids covered on either face by protein. The presence of protein on the surface supposedly makes the membrane *hydrophilic* rather than hydrophobic, meaning that water adheres easily to the membrane. The lipid center of the membrane supposedly makes the membrane impervious to lipid-insoluble substances. The small knobbed structures lying at the bases of the protein molecules are phospholipid molecules; the fat portion of the phospholipid molecule is dissolved in the lipid phase of the cell membrane, and the polar (ionized) portion protrudes outward where it is bound electrochemically with the protein that lines the outer surface of the membrane.

Pores in the membrane. The membrane has many minute pores, as shown in Figure 11B, that continue from one side to the other. These pores are so small that they have never been seen even with an electron microscope, but functional experiments to study the movement of substances between the extra- and intracellular fluids have demonstrated free diffusion of molecules up to a size of approximately 7 angstroms. It is through these pores that lipid-insoluble substances of very small sizes, such as water and urea molecules, pass with relative ease between the interior and exterior of the cell.

Formation of a new cell membrane—The "surface precipitation reaction." When the membrane of a cell becomes ruptured, cytoplasm streams out of the opening, but before it can flow far into the surrounding fluids, a *surface precipitation reaction* occurs; that is, fats and proteins precipitate to form a new membrane having exactly the same properties as the original membrane. This reaction will not occur when the concentration of calcium ions in the extracellular fluids is very low. Therefore, one of the very important functions of calcium ions is to maintain the integrity of the cell membrane itself. Also, the

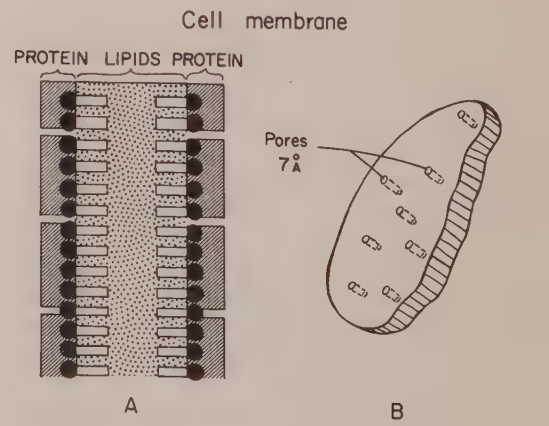


Figure 11. (A) Postulated molecular organization of the cell membrane. (B) Pores in the cell membrane.

degree of porosity of the membrane is determined by the concentration of the calcium ions—the greater the calcium ion concentration the less the porosity.

The Cytoplasm

Figure 12 illustrates the general structure of the cytoplasmic compartment of the cell. The cytoplasm is loaded with both minute and large dispersates, ranging in size from a few angstroms to one micron in diameter. The clear fluid of the cytoplasm in which the particles are dispersed is called the *hyaloplasm;* this contains mainly dissolved *protein enzymes, electrolytes, glucose,* and small quantities of *phospholipids, cholesterol,* and *esterified fatty acids.* The outer portion of the cytoplasm is frequently jelled into a semi-solid area called the *cortex* or *ectoplasm.* The fluid portion of the cytoplasm between the cortex and the nuclear membrane is called the *endoplasm.*

The most important organelles in the cytoplasm are the *endoplasmic reticulum, mitochondria, microsomes, lysosomes, secretory granules,* and the *Golgi complex.*

The endoplasmic reticulum. Figure 12 illustrates at the top a network of tubular and vesicular structures called the *endoplasmic reticulum.* The detailed structure of this organelle is shown in Figure 13. The space inside the tubules and vesicles is filled with *endoplasmic matrix,* a fluid medium that is different from the hyaloplasm on the outside of the endoplasmic reticulum.

The space inside the endoplasmic reticulum is

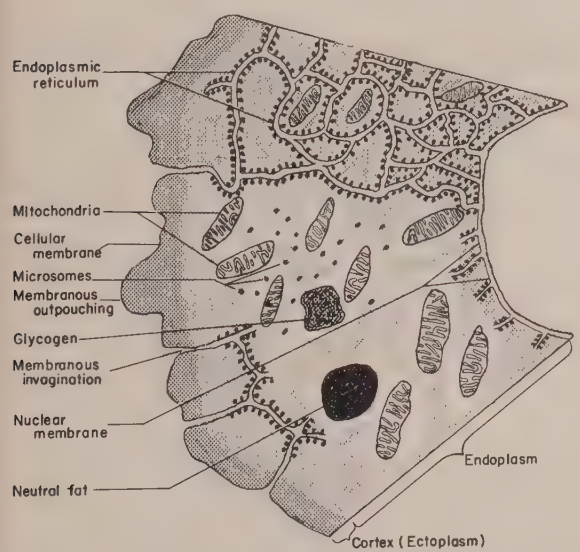


Figure 12. Organization of the cytoplasmic portion of the cell.

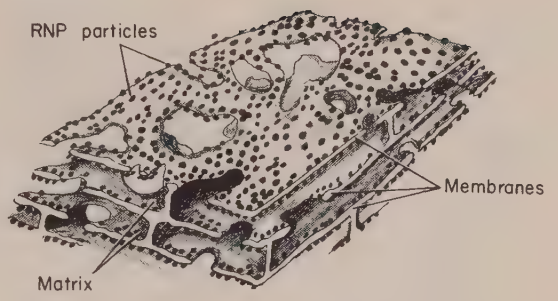


Figure 13. Structure of the membranes of the endoplasmic reticulum. (Redrawn from DeRobertis, Nowinski, and Saez: *General Cytology.* 3rd Ed.)

connected with a hollow space in the nuclear membrane and in some instances also with the exterior of the cell. It is presumed that the endoplasmic reticulum might provide a system for conducting substances out of the nucleus to all parts of the cytoplasm.

RIBOSOMES. Attached to the outer surfaces of many parts of the endoplasmic reticulum are large numbers of small granular particles called *ribosomes.* Where large numbers of ribosomes are attached, the reticulum is called either the *granular endoplasmic reticulum* or *ergastoplasm.* The ribosomes are composed mainly of *ribonucleic acid* which functions as an enzymatic template for the synthesis of protein in the cell, as we shall discuss in Chapter 4.

Part of the endoplasmic reticulum sometimes has no ribosomes. This is called the *agranular endoplasmic reticulum,* and it is believed to synthesize lipid substances and other chemicals necessary to cellular function.

Mitochondria. Mitochrondria are present in all cells, but the number varies from a few hundred to a few thousand, depending upon the amount of energy required by each cell to perform its functions. Figure 12 illustrates mitochondria dispersed throughout the cytoplasm. Mitochondria are also very variable in size and shape, some being only a few hundred millimicrons in diameter and globular in shape, while others are as large as one micron in diameter, as long as 7 microns, and filamentous in shape rather than globular.

The basic structure of the mitochondrion is illustrated in Figure 14, which shows that it is surrounded by a two-layered membrane probably similar in structure to that of the cell membrane. Many infoldings of the inner membrane form shelves on which almost all the oxidative enzymes of the cell are adsorbed. When nutrients and oxygen come into contact with these enzymes in the mitochondrion, they react to form carbon

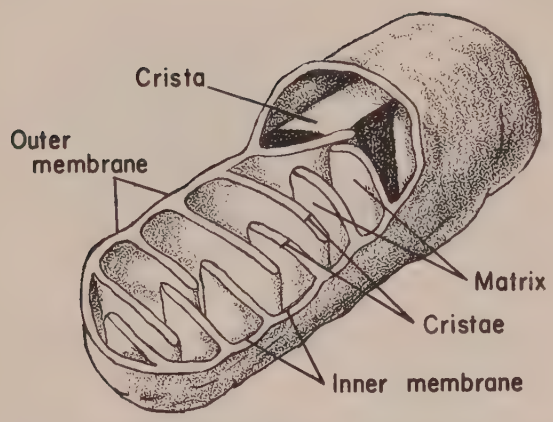


Figure 14. Structure of a mitochondrion. (Redrawn from Robertson: *Biochem. Soc. Symp.* 16:38. Cambridge University Press, 1959.)

dioxide and water, and the liberated energy is used to synthesize a substance called *adenosine triphosphate.* The adenosine triphosphate then diffuses throughout the cell and releases its stored energy wherever it is needed for performing cellular functions. The function of adenosine triphosphate is so important to the cell that we discuss it in detail in the next chapter.

Lysosomes. Another structure recently discovered in cells is the lysosome. Lysosomes are 250 to 750 millimicrons in diameter, and they, like the mitochondria, are surrounded by a lipoprotein membrane. They are filled with large numbers of small granules 55 to 80 angstroms in diameter which are protein aggregates of *hydrolytic enzymes.* A hydrolytic enzyme is capable of splitting an organic compound into two or more parts by combining the hydrogen portion of a water molecule with part of the compound and by combining the hydroxyl portion of the water molecule with the other part of the compound. For instance, protein is hydrolyzed to form amino acids, glycogen is hylrolyzed to form glucose, and neutral fats are hydrolyzed to form fatty acids and glycerol.

Ordinarily, the membrane surrounding the lysosomes prevents the enclosed hydrolytic enzymes from coming into contact with other substances in the cell. However, damage to the cell will break the membranes of some of the lysosomes, allowing release of hydrolytic enzymes. These then split the organic substances of the cell into small highly diffusible substances such as amino acids, glucose, fatty acids, and so forth. In this way, the damaged cell becomes dissoluted and often disappears from existence.

Microsomes. Located throughout the cytoplasm of all cells are large numbers of very minute granular particles varying in diameter from a few angstroms up to the sizes of lysosomes and mitochondria. These are grouped together under the term *microsomes,* which means simply very minute particles. A large portion of the microsomes are probably ribosomes that are not attached to the endoplasmic reticulum. These ribosomes, like those attached to the endoplasmic reticulum, are made up mainly of ribonucleic acid and presumably can synthesize protein in the same way that the ribosomes attached to the reticulum also synthesize protein. Other microsomes undoubtedly perform other functions that have not yet been discovered.

Secretory granules. One of the most important functions of many cells is secretion of special substances. The secretory substances are usually formed inside the cell and held there until an appropriate time for release to the exterior. The storage depots within the cells are called *secretory granules;* these are illustrated by the dark spots in Figure 15 which shows pancreatic acinar cells that have formed and stored enzymes that will later be secreted into the intestinal tract.

Many secretory granules lie inside the tubules and vesicles of the endoplasmic reticulum, though other secretory granules are probably free in the cytoplasm.

The Golgi complex. The Golgi complex is not found in all cells, and it has a very variable structure in different types of cells. Figure 16 illustrates a rather common structure of the Golgi complex located very near to the nucleus. The Golgi complex is usually highly lipid in nature, staining very darkly with dyes that have strong affinity for lipids; it probably contains a special secretory accumulation of lipid substances.

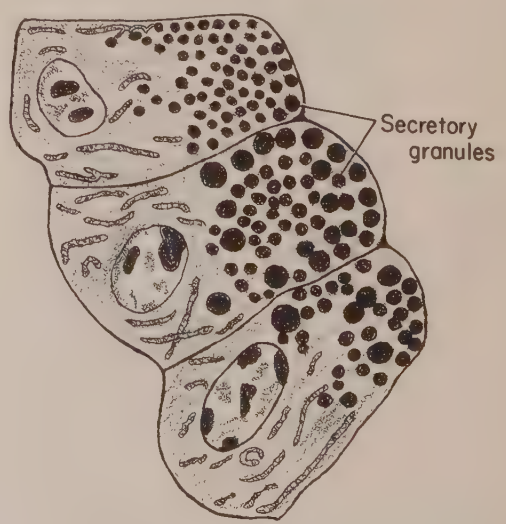


Figure 15. Secretory granules in acinar cells of the pancreas.

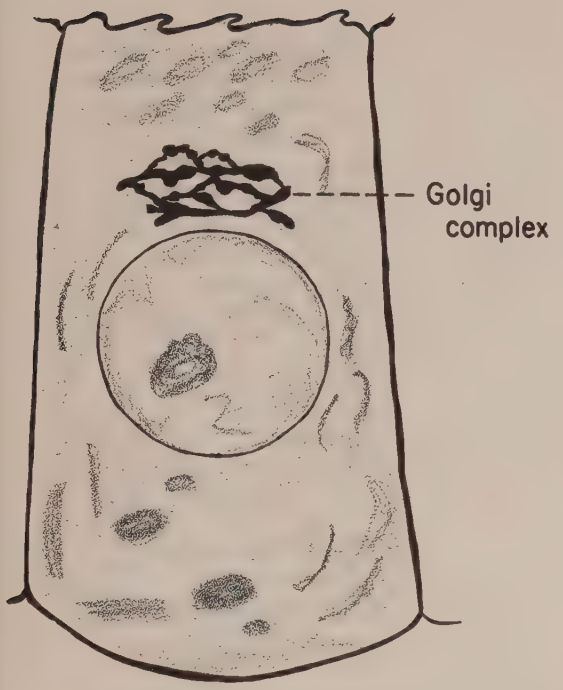


Figure 16. A typical Golgi complex.

Electron microscopic studies have shown the Golgi complex to be connected with the endoplasmic reticulum, indicating that substances synthesized in the endoplasmic reticulum might be shunted into large vesicles for storage, which then become what we call the Golgi complex. Therefore, in essence, the Golgi complex might be simply a specialized portion of the endoplasmic reticulum.

Other organelles of the cytoplasm. Besides the organelles described above, the cytoplasm of each cell contains two *centrioles,* small cylindrical structures that play a major role in cell division, as will be discussed in Chapter 4. Also, most cells contain small *lipid droplets* and *glycogen granules* which play important roles in energy metabolism of the cell. And certain cells contain highly specialized structures such as the *cilia* of ciliated cells which are actually outgrowths of the cytoplasm, and the *myofibrils* of muscle cells; these will be discussed in detail at different points in this text.

The Nucleus

The nucleus is the control center of the cell. It controls the chemical reactions that occur in the cell and also reproduction of the cell. The precise means by which it effects these controls is still partially obscure. Briefly, the nucleus contains large quantities of *deoxyribonucleic acid,* which is what we have called *genes* for many years. The genes control the characteristics of the protein enzymes of the cytoplasm and in this way control cytoplasmic activities. To control reproduction, the genes autocatalytically reproduce themselves, and after this is accomplished, the cell splits by a special process called *mitosis* to form two daughter cells, each of which receives one of the two sets of genes. These reproductive activities of the nucleus will be considered in detail in Chapter 4.

The structure of the nucleus does not give us much clue to the mechanism by which the nucleus performs its control activities. Figure 17 illustrates the *interphase nucleus* (period between mitosis), showing a finely granular dispersion of darkly staining *chromatin material* throughout the fluid of the nucleus called the *nuclear sap.* The chromatin material is mainly deoxyribonucleic acid and, therefore, is the genes. During the period of mitosis, as will be discussed in Chapter 4, this chromatin material aggregates to form the highly structured *chromosomes* which can be seen easily with a microscope. It is presumed that even during the interphase period of cellular activity the granular chromatin material might still be organized into chromosomal structures by some means not yet discovered.

The nucleolus. The nuclei of many cells contain a lightly staining structure called the *nucleolus.* Unlike most of the organelles that we have discussed, this structure does not have a limiting membrane. Instead, it is simply an aggregate of loosely bound granules composed mainly of *ribonucleic acid.* The nucleolus usually becomes considerably enlarged when a cell is actively synthesizing other proteins. It is believed that the genes synthesize the ribonucleic acid and store it in the nucleolus; then this ribonucleic acid later disperses from the nucleolus into the cytoplasm where it controls cytoplasmic function. The

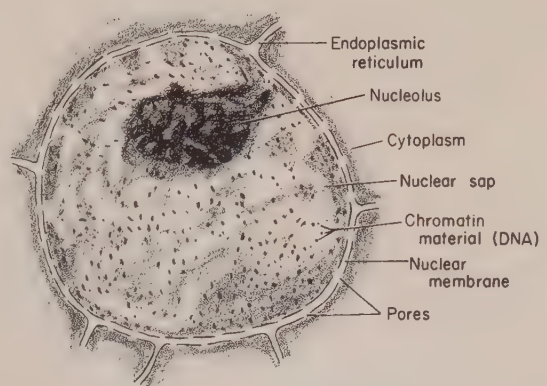


Figure 17. Ultrastructure of the nucleus.

details of this mechanism will be discussed in Chapter 4.

The nuclear membrane. The nuclear membrane is a very porous two-layered structure. Electron micrographs show the sizes of its pores to be 400 to 700 angstroms, large enough to allow even the largest protein molecules to pass relatively freely between the nucleus and the cytoplasm. Thus, it is believed that even the large ribonucleic acid molecules synthesized by the genes in the nucleus pass out of the nucleus into the cytoplasm where they exert their control over cytoplasmic chemical reactions.

Electron microscopic studies have shown direct connections between the nuclear membrane and the endoplasmic reticulum, showing that the lumens of the reticulum interconnect with the space between the two layers of the nuclear membrane. It has been postulated that some substances formed in the nucleus enter the space in the nuclear membrane and then flow into the endoplasmic reticulum, in this way passing to all parts of the cytoplasmic cavity.

Colloidal Nature of Protoplasm

This chapter would not be complete without pointing out the *colloidal* properties of both cytoplasm and nucleoplasm, for in both of these many finely dispersed particles float freely in the fluid—for instance, the *microsomes, lysosomes,* and *mitochondria* in the cytoplasm and the *chromatin* in the nucleoplasm. The particles are *hydrophilic*—that is, attracted to the water—because of polar charges on their surfaces, and they remain *dispersed* because of *mutual repulsion* of these charges. Finally, the particles exhibit other properties of colloids such as the *scattering of light* as it passes through the solution, which gives protoplasm a milky appearance, and such as its *ability to gel* under certain circumstances.

The very high concentrations of proteins in protoplasm help to give it its colloidal properties, for large protein molecules have some properties of colloids. Also, the protein molecules tend to polymerize into aggregates which then become true colloids.

REFERENCES

Brachet, J.: *Biochemical Cytology.* New York, Academic Press, 1957.

Brachet, J.: *The Cell.* Volume I. New York, Academic Press, 1958.

Engstrom, A., and Finean, J. B.: *Biological Ultrastructure.* New York, Academic Press, 1958.

Giese, A. C.: *Cell Physiology.* 2nd Ed. Philadelphia, W. B. Saunders Company, 1962.

McElroy, W. D.: *Cellular Physiology and Biochemistry.* Englewood Cliffs, Prentice-Hall, Inc., 1960.

Symposium on Biochemical Aspects of Cell Structure and Function. *Fed. Proc.* 18:957, 1959.

Walker, P. M. B., editor: *New Approaches in Cell Biology.* New York, Academic Press, 1960.

BASIC FUNCTIONAL SYSTEMS OF THE CELL

3

In this chapter we will discuss several of the major functional systems of the cell, particularly transport of substances through the cell membrane, energy processes in the cell, synthesis of intracellular substances, and functions characteristic of certain special cells, such as ameboid motion and ciliary activity.

TRANSPORT OF SUBSTANCES THROUGH THE CELL MEMBRANE

Each cell receives its nutrients from the surrounding fluids and in turn empties the end-products of its metabolic reactions back into these same fluids. The nutrients and end-products are transported through the cell membrane in one of three ways, by *diffusion,* by *active transport,* or by *pinocytosis.*

Transport by Diffusion

The term *diffusion* means random motion of all molecules and ions in all directions, which results from kinetic energy stored in all these particles, making them move continually, bouncing against each other, each particle weaving its way among all the others, and continuing to go wherever it is not limited by some solid structure. This process can be likened to dancers on a dance floor, the motion continuing in all directions.

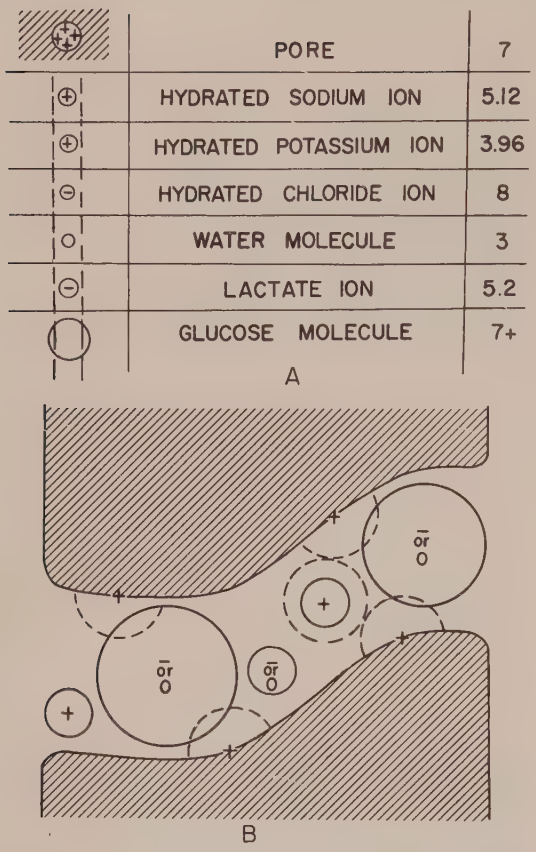


Figure 18. (A) Sizes (in microns) and charges of different molecules and ions in relation to the cellular pore size. (B) Postulated nature of the cellular pore, showing sphere of influence exerted by positive charges along the surface of the pore. (Modified from Solomon: *Scientific American,* Dec., 1960, p. 146.)

Substances can be transported through the cell membrane by diffusion either through the *pores* or through the *matrix* of the membrane itself.

Diffusion through the pores. The pores in membranes have never been seen, even with the electron microscope, because they have a diameter of only 7 angstroms, which is smaller than the limit of resolution by this apparatus. Yet, physiological experiments, in which the passages of different size molecules through the membrane have been studied, have characterized the size of the pores, the number of pores, and other special characteristics that determine whether or not molecules and ions will diffuse through them. The total area of the pores represents only *one sixteen-hundredth* of the total cell surface area, the remainder being unbroken cell membrane having a lipoprotein matrix which was described in the previous chapter.

Figure 18A illustrates the size of the membrane pore in relation to the sizes of common molecules and ions in the cellular fluids. This figure shows that the water molecule is less than one-half the size of the pore, which allows extremely rapid diffusion of water. It shows also that potassium, sodium, chloride, and lactate ions are all much smaller than the pore. On the other hand, the glucose molecule is slightly larger than the pore, which means that glucose cannot diffuse through the usual pore, though it might diffuse through the rare oversized pore. Experiments on unicellular animals have shown that water diffuses through their pores about 40,000 times as easily as glucose, illustrating that, from a practical standpoint, glucose is essentially excluded from passage through the membrane by this route.

DIFFUSION OF WATER AND OTHER NONCHARGED MOLECULES. The most important substance that diffuses through the pores of the cell membrane is water. Water is not soluble in the matrix of the membrane and, therefore, cannot cross the cell membrane in any other way except through the pores. Even though only an extremely small portion of the cell membrane is represented by pores, the amount of water that diffuses through the pores each second is equal to approximately 100 times the volume of the cell itself. This amazing rate of transfer illustrates that despite the small number of pores the process of diffusion is so rapid that extreme amounts of water can enter or leave a cell each second. Other *noncharged molecules* can also diffuse through the cell membrane in the same way if their sizes are small enough for them to go through the pores.

DIFFUSION OF IONS. *Negatively charged ions* pass through cell pores in the same way as water molecules, but *positively charged ions* have very great difficulty in passing through. For instance, potassium ions, which are positively charged, have almost one million times as much difficulty in passing through the pores as do the negatively charged chloride ions, even though their particle diameters are almost exactly the same. To explain this difference, the pore structure has been proposed to be that shown in Figure 18B. This figure shows positive and negative ions and uncharged molecules passing through the 7-angstrom pore, but it also shows the surfaces of the pore lined with positively charged substances. These positive charges exert a "sphere of influence" extending into the pore which repels positive ions. Therefore, positive ions, in attempting to diffuse through the pores are generally repelled back out of the pore before they can go through.

Diffusion through the matrix of the membrane. In the previous chapter, it was pointed out that the matrix of the cell membrane has a lipoprotein

structure and that the central portion of the membrane is probably composed almost entirely of lipids. For this reason, lipid-insoluble substances cannot diffuse through the matrix of the membrane, but a few substances important to the function of the cell are *lipid-soluble,* such as *oxygen, carbon dioxide,* and *alcohol.* Therefore, these molecules can pass through the membrane by becoming dissolved in the matrix and then diffusing to the other side.

FACILITATED DIFFUSION. Some substances that are not lipid-soluble can diffuse through the matrix of the membrane by a process called *facilitated diffusion,* illustrated in Figure 19A. The substance on coming in contact with the surface of the membrane combines with a *carrier,* which is a chemical compound present in the membrane itself. Even though the substance to be transported is not soluble in the membrane, the combination of carrier and substance *is* soluble, and this combination diffuses to the inner surface of the membrane. At this point, the substance breaks away from the carrier and is released to the interior of the cell. In turn, the carrier diffuses back through the membrane to its outer surface where it picks up additional amounts of the substance.

The most important substance that normally passes through the cell membrane by the process of facilitated diffusion is *glucose,* which is neither small enough to go through the pores nor lipid-soluble, and, yet, is needed in great amounts by the cells for energy.

A particular value of facilitated diffusion is that it selectively controls the types of substances that enter the cell. Such selection can occur in two ways: First, each type of carrier in the membrane is selective for the material that it will transport. Second, the chemical combination of the carrier with each transported substance is catalyzed by *specific enzymes;* therefore, only those substances for which specific enzymes and carriers are available can be transported in this way.

The "diffusion gradient." Transport by the process of diffusion, whether it be by facilitated diffusion or by simple diffusion, occurs only when the concentration on one side of the membrane is greater than on the other side. Obviously, if the concentrations are the same on the two sides, random motion will make as many molecules move in one direction as in the other so that no *net* transport will occur. But, if the concentrations of the two sides are not the same, more molecules will move from the side of higher concentration to the other side. Furthermore, the quantity that is transported each minute *is directly proportional to the difference between the concentrations on the two sides of the membrane,* which is called the *diffusion gradient.*

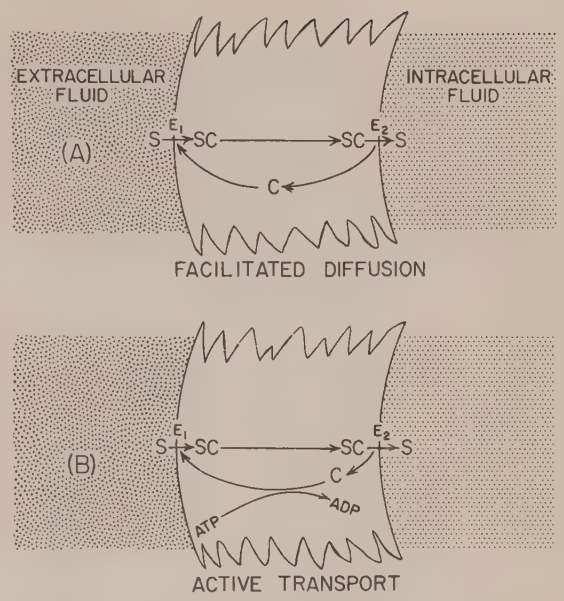


Figure 19. Dynamics of (A) facilitated diffusion and (B) active transport through the cellular membrane. C represents carrier, S represents a substance passing through the membrane, and E represents enzymes.

Active Transport

Active transport, like facilitated diffusion, also involves transport of substances through the membrane by carriers. Yet it is different from facilitated diffusion in two major ways: First, active transport can cause substances to cross the membrane even when the concentration of the substance is higher on the side of the membrane where the substance is going than on the side from which it is coming. Second, energy is expended in causing active transport against this *concentration gradient.*

Figure 19B shows the basic mechanism of active transport and also illustrates its difference from facilitated diffusion, showing that for transport to occur energy must be released by the breakdown of adenosine triphosphate (ATP) to adenosine diphosphate (ADP), a process that is discussed in greater detail later in the chapter. The chemical reactions for breakdown of ATP to ADP are *coupled* with the chemical reactions that cause active transport, and energy released from the ATP is transferred to the substance being transported, causing its movement from the area of dilute concentration to the area of increased concentration.

Sodium and *potassium ions* are actively transported through the membranes of all cells in the body, and other substances actively transported by the membranes of some cells include *chloride ions, hydrogen ions, phosphate ions, calcium ions, magnesium ions, creatinine,* and *uric acid.*

Active transport of sodium and potassium is a major factor in maintaining the normal differences in electrolytic compositions of the fluids inside and outside the cell. For instance, the normal concentration of potassium inside a cell of the human body is about 130 milliequivalents per liter while that outside is about 4 milliequivalents per liter. The sodium concentrations are almost exactly opposite to these, the concentration outside the cell being 140 milliequivalents per liter and inside 10 milliequivalents per liter. Thus, the active transport mechanisms for these two substances operate in opposite directions, sodium being transported out of the cell and potassium being transported into the cell.

The sodium transport mechanism creates the electrical potentials that are necessary for transmission of impulses in both nerve and muscle fibers. Therefore, the sodium transport process has been studied more than any of the other transport mechanisms and has been called the "sodium pump." Thus far, the specific enzymes necessary for active transport of sodium have not been isolated, but the carrier is believed to be *phosphatidic acid* which is a combination of glycerol, two fatty acid molecules, and phosphoric acid. It is a type of phospholipid which, like other lipids, is soluble in the matrix of the cell membrane. It also has two negative valences which allow it to combine readily with sodium.

Pinocytosis

Pinocytosis is a process by which small quantities of surrounding fluid are imbibed into the cell by invagination of the cell membrane. Figure 20 illustrates this process. Figure 20A shows substances in the extracellular fluid that cannot enter the cell by either diffusion or active transport, such as *protein molecules* or *lipoproteins.* Figure 20B illustrates these substances becoming adsorbed to the membrane, and Figure 20C shows this causing the surface of the membrane to invaginate. Then the membrane closes over, as illustrated in Figure 20E. Finally, the vesicle disintegrates, and its contents become dispersed into the cytoplasm as shown in Figure 20F.

Pinocytosis takes place only in response to certain substances. Proteins, in particular, or gross changes in the electrolyte composition of the

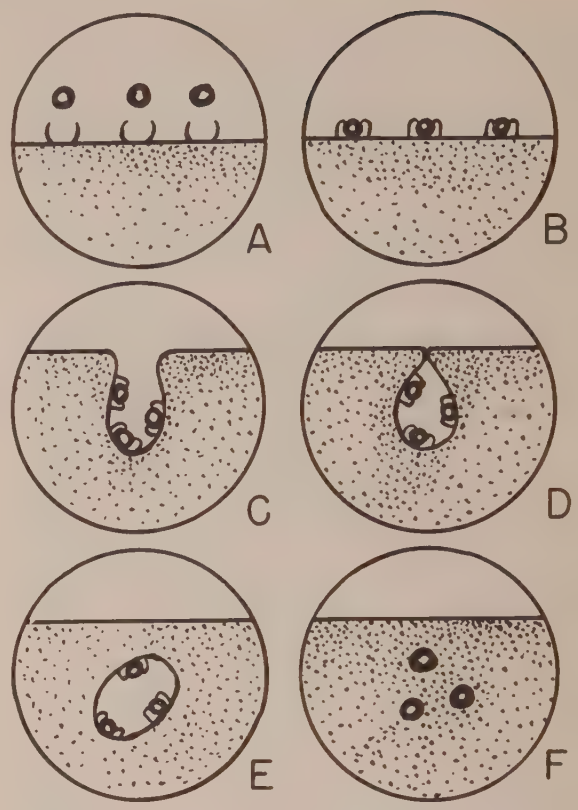


Figure 20. Mechanism of pinocytosis. (From Bennett: *J. Biophys. Biochem. Cytol.* 2: part 2 (Suppl.), 99, 1956.)

fluids promote this process. However, regardless of what initiates pinocytosis, a considerable amount of the extracellular fluid and its dissolved constituents becomes included in the pinocytotic vesicle and therefore enters the cell along with the substance actually initiating the process.

Pinocytosis is particularly important for transporting large molecular weight substances, especially proteins, through the cell membrane, because there are no other processes by which these substances can be transported. Some examples of pinocytosis in the human being are: (1) transport of protein out of the tubules of the kidneys into the tubular epithelium, (2) transport of plasma into the endothelial cells lining the capillaries, (3) probably transport of protein hormones into cells, and (4) probably transport of albumin and globulin into almost all cells of the body to be used as nutrients.

The process of pinocytosis is extremely slow in comparison with direct diffusion of small molecules through the cell membrane. For instance, water *diffuses* through the cell membrane about one million times as rapidly as it can be moved into the cell by pinocytosis. Even so, this slow transport by pinocytosis is still extremely impor-

tant to the cell, for it is the *only* method for transport of some very large substances.

ENERGY METABOLISM

Formation of ATP

The basic energy system of all cells is simply the following: (1) Each cell has enzymes in the lysosomes for splitting the three major foods of the diet—*carbohydrates, fats,* and *proteins*—into small chemical compounds, *glucose, fatty acids,* and *amino acids* respectively. (2) These small compounds are then metabolized with oxygen to form carbon dioxide and water, at the same time releasing tremendous quantities of energy. (3) This energy is used almost entirely for a single purpose, to convert *adenosine diphosphate* (*ADP*) into *adenosine triphosphate* (*ATP*). (4) The ATP then supplies the needed energy to the different chemical reactions in the cell.

Figure 21 illustrates the release of energy from fatty acids, glucose, and amino acids. Glucose is converted first to *pyruvate* (pyruvic acid) in the cytoplasm, while both the fatty acids and amino acids are converted to *acetoacetic acid.* Then the pyruvate and the acetoacetic acid are converted to the same compound, *acetyl coenzyme A,* which enters the mitochondria. In the mitochondria the acetyl coenzyme A is split into a succession of smaller compounds, with the hydrogen and carbon atoms being progressively removed and combined with oxygen to form carbon dioxide and water.

About 5 per cent of the ATP synthesized in the cell is formed in the initial stages of these chemical reactions, that is, while the glucose, fatty acids, and amino acids are being converted to acetyl coenzyme A. About 95 per cent of the ATP is formed as the acetyl coenzyme A is converted into carbon dioxide and water in the mitochondria. Because of the extreme amount of "energy-rich" ATP formed in the mitochondria, the mitochondria are frequently called the "powerhouses" of the cell.

Anaerobic and aerobic energy metabolism. The formation of adenosine triphosphate in the initial stages of nutrient breakdown—that is, while the glucose, fatty acids, and amino acids are being converted to acetyl coenzyme A—does not require oxygen. Therefore, this phase of energy metabolism is called the *anaerobic phase.* On the other hand, the formation of ATP inside the mitochondrion will occur only in the presence of oxygen. Therefore, the energy metabolism inside the mitochondrion is called the *aerobic phase* of energy metabolism. Since only a small quantity of adenosine triphosphate is formed during the anaerobic phase of metabolism, only a small amount of energy can be supplied to the cell without oxygen, and the end-products of anaerobic metabolism—pyruvic acid, acetoacetic acid, and acetyl coenzyme A—become so concentrated in the cell within a minute or so that even anaerobic metabolism ceases. Therefore, oxygen is absolutely necessary if any significant degree of energy metabolism is to continue for longer than a minute or two at a time.

Cellular respiration. Figure 21 also shows the uptake of oxygen by the cells and release of carbon dioxide. Oxygen diffuses through the cell membrane into the cytoplasm and thence through

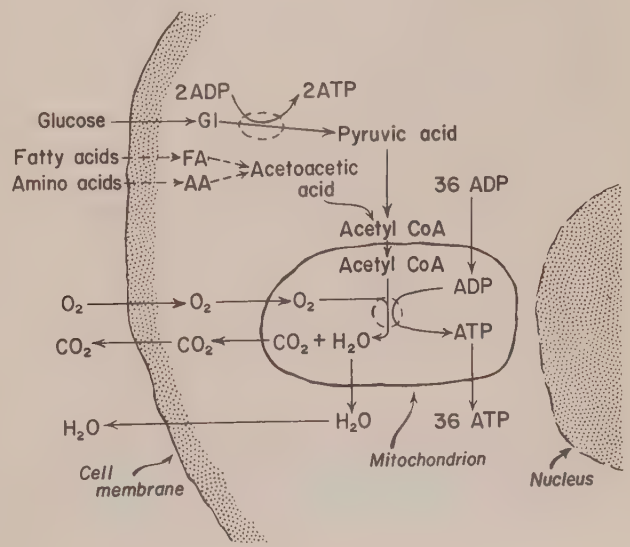
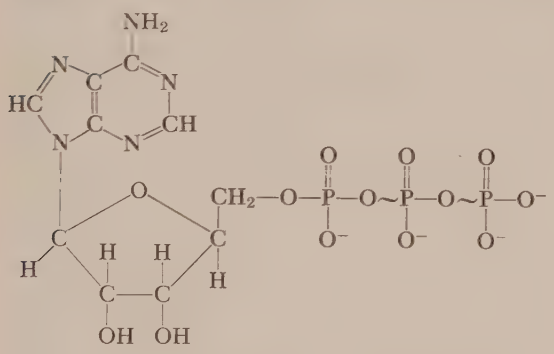


Figure 21. Formation of adenosine triphosphate in the cell, showing that most of the ATP is formed in the mitochondria.

the mitochondrial membrane into the matrix of the mitochrondrion. It is here that essentially all oxygen usage occurs in the cell, the oxygen combining with the split products of acetyl coenzyme A. In turn, large quantities of carbon dioxide are continually formed in the mitochondria. As a result, it diffuses outward into the cytoplasm, thence through the cell membrane into the surrounding fluids. It is fortunate that both oxygen and carbon dioxide are reasonably soluble in the lipid substances of the cellular and mitochondrial membranes, for this allows extremely rapid passage of both of these molecules between the extracellular, intracellular, and intramitochondrial fluids.

The special properties of adenosine triphosphate. The adenosine triphosphate molecule is a nucleotide composed of the nitrogenous base *adenine*, the pentose sugar *ribose*, and three molecules of *phosphoric acid*, the formula for which is:



The last two phosphoric acid radicals are connected with the remainder of the molecule by means of so-called *high energy phosphate bonds*, which are represented by the symbol ~. Each of these bonds contains about 7000 calories of energy per mole of adenosine triphosphate, which is many times the energy stored in the average chemical bond, thus giving rise to the term "high energy" bond. Furthermore, the high energy phosphate bonds are very labile so that they can be split very rapidly on demand wherever in the cell energy is required to promote other chemical reactions. Thus, adenosine triphosphate is an almost "explosive-like" compound carrying large amounts of energy, a compound that diffuses from its anaerobic and aerobic points of formation in the cell to all other parts of the cell to be used rapidly and easily any place where energy is required.

When adenosine triphosphate releases its energy, a *phosphoric acid radical* is split away, and *adenosine diphosphate* (ADP) is formed. These two substances are then recombined by the anaerobic and aerobic processes to form new adenosine triphosphate, the entire process continuing over and over again. For these reasons, adenosine triphosphate has been called the *energy currency* of the cell, for it can be "spent" and remade again and again.

Uses of adenosine triphosphate for cellular function. Adenosine triphosphate is used to promote three major categories of cellular functions: (1) *membrane transport*, (2) *synthesis of chemical compounds* throughout the cell, and (3) *performance of mechanical work*. These three different uses of adenosine triphosphate are illustrated in Figure 22 which shows adenosine triphosphate being used (a) to supply energy for the transport of glucose through the membrane, (b) to promote protein synthesis by the ribosomes, and (c) to supply the energy needed during muscle contraction.

In addition to membrane transport of glucose, energy from adenosine triphosphate is required for transport of sodium ions, potassium ions, and, in certain cells, calcium ions, phosphate ions, chloride ions, uric acid, hydrogen ions, and still many other special substances. Membrane transport is so important to cellular function that some cells utilize as much as 30 per cent of the adenosine triphosphate for this purpose alone.

In addition to synthesizing proteins, cells also synthesize phospholipids, cholesterol, purines, pyrimidines, and a host of other substances. Synthesis of almost any chemical compound requires energy. For instance, a single protein molecule might be composed of as many as several thousand amino acids attached to each other by peptide linkages, and the formation of each one of these linkages requires the breakdown of an adenosine triphosphate molecule to adenosine

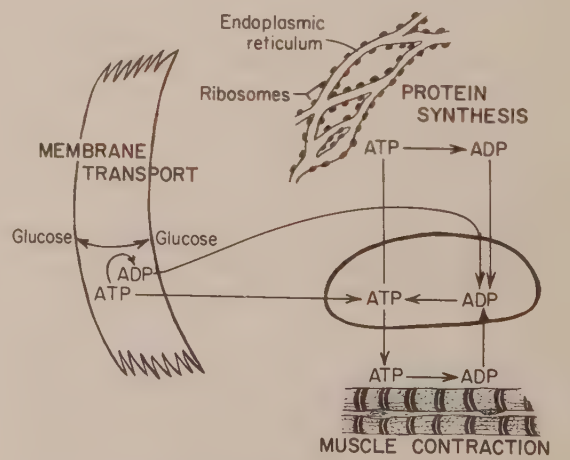


Figure 22. Use of adenosine triphosphate to provide energy for three of the major cellular functions, (1) membrane transport, (2) protein synthesis, and (3) muscle contraction.

diphosphate; thus several thousand ATP molecules must release their energy as each protein molecule is formed. Indeed, cells often utilize as much as 75 per cent of the adenosine triphosphate formed in the cell simply to synthesize new chemical compounds; this is particularly true during the growing phase of cells.

The final major use of adenosine triphosphate is to supply energy for special cells to perform mechanical work. We shall see in Chapter 21 that each contraction of a muscle fibril requires expenditure of tremendous quantities of adenosine triphosphate. This is true whether the fibril is in skeletal muscle, smooth muscle, or cardiac muscle. Other cells perform mechanical work in two additional ways, by *ciliary motion* or by *ameboid motion*, both of which are discussed later in this chapter. The source of energy for all these types of mechanical work is adenosine triphosphate.

In summary, adenosine triphosphate is always available to release its energy rapidly and almost explosively wherever in the cell it is needed. To replace the adenosine triphosphate used by the cell, much slower chemical reactions break down carbohydrates, fats, and proteins, and use the energy derived from these to form new adenosine triphosphate.

SYNTHESIS OF PROTEIN AND OTHER CELLULAR SUBSTANCES

Proteins

In the previous chapter it was pointed out that all cells have two different types of proteins, *structural proteins* and *enzymes*. The structural proteins are a necessary part of all the cellular membranes and other physical structures of the cell. On the other hand, the enzymes control the chemical reactions in the cell. Therefore, the types of proteins in the cell determine both the cell's structure and its functions.

Proteins are synthesized mainly by the *ribosomes*, which are small particulate structures attached to the outside of most of the endoplasmic reticulum but which probably sometimes float freely in the cytoplasm as well. The ribosomes in turn are composed mainly of *ribonucleic acids* (RNA) which are combined with proteins to form *ribonucleoproteins*. The types of ribonucleic acids in the ribosomes are controlled by similar substances called *deoxyribonucleic acids*, which are the genes in the nucleus. Therefore, indirectly, the genes in the nucleus control the characteristics of the proteins produced in the cell, which in turn control the character of the cell itself. This subject is discussed in detail in the following chapter.

Another type of substance that is essential for protein synthesis is the *soluble* ribonucleic acids that are dissolved in the cytoplasm. These nucleic acid molecules are much smaller than the ribosomal nucleic acid molecules, and their functions are also different, as we shall see below. Twenty-three different types of soluble ribonucleic acids are present in the cytoplasm, each one of which combines specifically with one of the 23 different types of amino acids.

The steps for synthesis of the protein molecule are shown in Figure 23. At the top are shown three amino acids designated respectively AA_1, AA_2, and AA_{20}. The same reactions that these three amino acids go through apply for all the other 20 amino acids. The first stage in the formation of a protein molecule is *activation* of the *amino acids*, which means transfer of a large quantity of energy to them. In this process, adenosine triphosphate reacts with the amino acids to form a complex composed of adenosine monophosphate and amino acid. In the next stage, the activated amino acid *combines with the specific soluble RNA* for that respective type of amino acid, and at the same time the adenosine monophosphate is removed. This same reaction occurs for each of the 23 amino acids. These combinations of soluble RNA's and their respective amino acids now come in contact with the ribosomal RNA which is a long, fibrillar polymer of many smaller RNA nucleotides. The different purine and pyrimidine bases in each of these RNA's are believed to be arranged in a definite pattern, the pattern constituting a "key" to the type of protein that is to be produced. The key determines the points along the fibril where each of the 23 specific soluble RNA molecules will attach to the ribosomal RNA as shown in Figure 23. In this way, the different amino acids are lined up in the appropriate order to form the predetermined type of protein molecule. Energy to cause attachment of the soluble RNA's with the ribosomal RNA is provided by *guanine triphosphate*, a substance very similar to adenosine triphosphate.

After the respective soluble RNA's are attached to the ribosomal RNA, the amino acid molecules have then become lined up in the pattern specified by the key in the ribosomal RNA molecule. There may be as few as 50 such amino acids lined up or as many as several thousand. Once this lining up process has taken place, the activated amino acids, having a large amount of stored energy because of the original activation process, automatically combine with each other

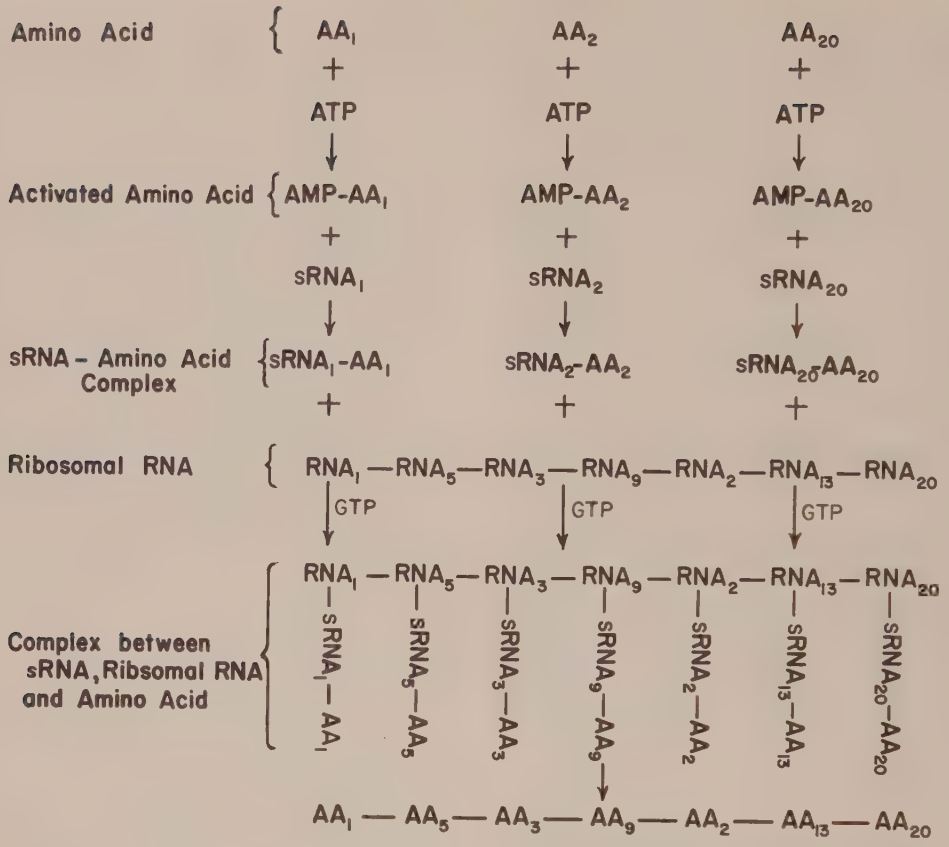
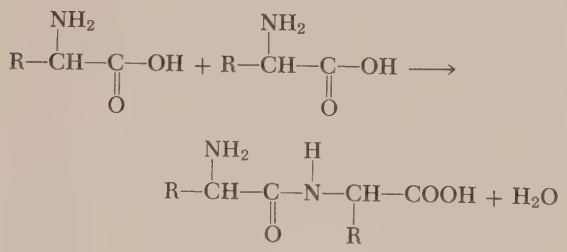


Figure 23. Sequential steps in the formation of a protein molecule, showing the following stages: (a) activation of the different amino acids, (b) combination of the activated amino acids with soluble RNA, (c) attachment of the soluble RNA–amino acid complex with ribosomal RNA, and (d) formation of peptide linkages and separation of the protein molecule from the complex.

to form peptide linkages according to the following typical reaction:



The peptide linkage of the amino acids is believed to occur by a "zipper" procedure. That is, the first two amino acids at one end of the molecule link together, then the next two, the next two, and so forth, the linkages occurring at a rate of several every second. Thus, a completed protein molecule is believed to be formed every minute or so by each ribosome during the growth phase of the cell. As the peptide linkages take place, the amino acids separate from the soluble RNA's, and the soluble RNA's in turn separate from the ribosomal RNA. These RNA's then perform these same functions again and again to synthesize successive identical protein molecules.

Synthesis of Other Compounds in the Cell

The synthesis of most other compounds in the cells is controlled by the protein enzymes, each one of the protein enzymes catalyzing one type of chemical reaction. The sites of most of these syntheses are unknown. It is possible that large quantities of substances are synthesized in the fluid of the cytoplasm itself. However, it is also believed that the agranular portion of the endoplasmic reticulum is a site for adsorption of large quantities of protein enzymes that promote formation of many non-protein molecules. Particularly, has the agranular endoplasmic reticulum been implicated in the formation of lipid substances.

Thus, in essence, the genetic key located in the nucleus controls the formation of the ribosomal key in the cytoplasm, and this controls the formation of the cytoplasmic proteins. These then control all the other chemical processes of the cell including synthesis of non-protein compounds.

Catalysis of Cellular Substances—Function of the Lysosomes

Aside from the formation of cellular substances, there also come times when it is important to digest substances in the cell or even times when it is important to dissolute the entire cell. These are functions of the *lysosomes,* which, as described in the previous chapter, are cytoplasmic organelles having a limiting membrane and filled with *hydrolytic protein enzymes.* Whenever these membranes are destroyed, the hydrolytic enzymes are released into the cytoplasm where they begin "hydrolyzing" the different cellular chemicals.

Most large organic compounds in cells are bound together by the process of *condensation* which is the opposite of hydrolysis. Condensation means combination of two molecules with the release of water, while hydrolysis means splitting a molecule into two molecules by addition of water, the hydrogen ion from the water molecule going to one of the newly formed molecules and a hydroxyl ion going to the other newly formed molecule. Typical cellular compounds that are formed by the process of condensation include the neutral fats, glycogen, and the proteins (the peptide linkage, for instance, is a condensation process).

When a cell becomes severely damaged, the membranes of the lysosomes rupture, releasing hydrolytic enzymes into the cytoplasm. The enzymes set to work splitting the proteins into amino acids, glycogen into glucose, and neutral fats into fatty acids and glycerol. In addition, many other compounds are similarly split into small molecular particles. Thus, the hydrolytic enzymes dissolute either part of the cytoplasm or the entire cell. In this way damaged parts of cells or entire cells are removed. Also, substances that might enter the cell, such as foreign proteins, whole bacteria or nutrients, can be dissoluted in the same way. Once dissolution has occurred, the small molecules of amino acids, glucose, fatty acids, and so forth can then be used by other parts of the cell for synthetic processes, or they can be extruded from the cell and then used by other cells.

SPECIAL CELLULAR FUNCTIONS

At many points in this text, we will discuss the functions of many special cells in the different organs. For instance, contraction of the muscle cells will be discussed in Chapter 21, transmission of nerve impulses in Chapter 20, reception of light by the rods and cones of the eye in Chapter 27, and phagocytosis by white blood cells in Chapter 9. However, two particular cellular functions deserve mention here because appropriate opportunity to discuss these along with other systems will not occur. These are the specialized motor activities of ameboid motion and ciliary activity.

Ameboid motion. Ameboid motion, named for *amebae* which exhibit well this type of movement, is illustrated in Figure 24. Its basic mechanism is the following: A *pseudopodium* projects forward and makes contact with a new point on the surface along which the cell is moving. Then the pseudopodium contracts, pulling the cell forward with either a rolling or a progressive arching movement. A similar type of movement can occur through the tissue spaces, this movement being exhibited particularly by white blood cells and tissue histiocytes.

Ameboid motion probably results from progressive "solation" and "gelation" of the cytoplasm in the cortex of the cell as illustrated in Figure 24. Some chemical agents on coming in contact with one side of the cell cause solation of the adjacent cell cortex; that is, cause the cortex to change from the gel state to the soluble state. This decreases the elastic tension of the cell membrane and allows this portion of the cell to protrude, forming the pseudopodium. Immediately thereafter, the cortex gels again, causing the membrane of the pseudopodium to become strong once more and to contract, pulling the cell forward. Additional sequences of solation and gelation occur repeatedly as long as the agent is present, and the cell continues to move toward the chemical stimulus; this phenomenon is known as *positive chemotaxis.* Occasionally, other substances cause cells to move away from the chemical stimulus, which is known as *negative chemotaxis.*

A very important instance of positive chemotaxis is the movement of white blood cells into infected areas of the tissues. The infected areas release a substance called *leukotaxine* that diffuses through the tissues and acts on the white blood cells to cause positive chemotaxis toward the area of origin of the leukotaxine. Thus, large numbers of white blood cells move into the

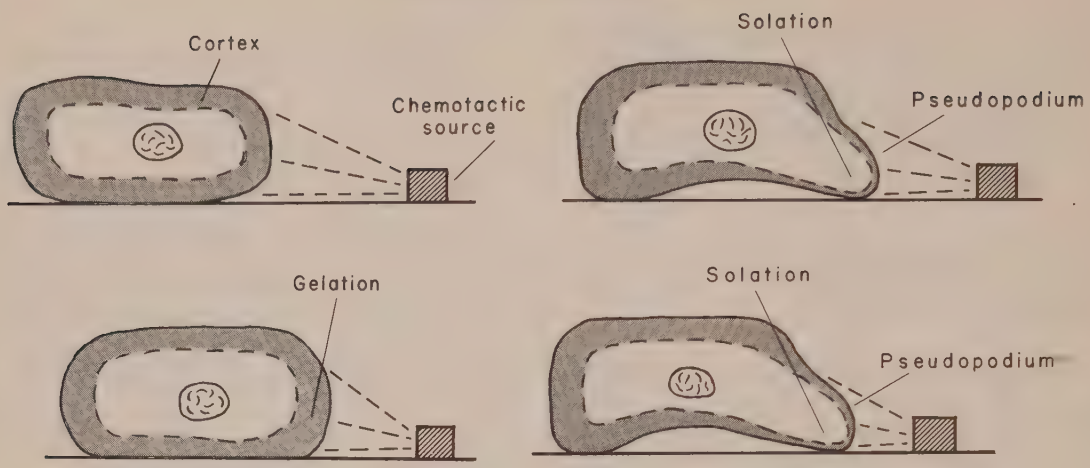


Figure 24. Mechanism of ameboid motion toward a chemotactic source.

infected area and help to destroy the infectious agent.

Ciliary motion. A second type of cellular motility that occurs on certain epithelial surfaces is *ciliary movement.* The *cilia* of a cell, as illustrated in Figure 25, are very minute fibrillar structures protruding from the cell's surface. Under the microscope these look to be minute, sharp-pointed hairs which sometimes project 3 to 4 microns away from the cell.

Cilia move forward very slowly and then backward suddenly, this occurring repetitively and continually. Because the resistance to movement of an object through a fluid medium is proportional to the square of its velocity of movement, the resistance to the forward slow movement of the cilium is very slight but to the backward rapid movement very great. Therefore, fluids on the cell surface are propelled in the direction of the rapid component of movement, and because the cilia on the surface of a ciliated epithelium are all oriented in the same direction, fluids are continually moved in one direction along the surface. It is in this way that secretions in the bronchi and trachea are continually moved upward along the ciliated respiratory epithelium toward the mouth.

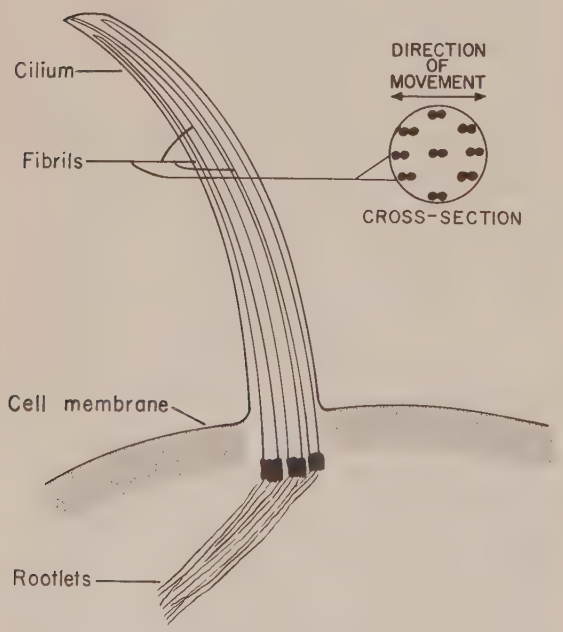


Figure 25. Typical structure of a cilium.

The mechanism that causes ciliary movement is probably the same for all cilia, whether they be found in uniciliated organisms or in the multiciliated epithelial cells in the human body, for electron microscopic studies of cilia have shown that each cilium of all types contains a pair of central filaments and a number of other paired filaments arranged radially around the border of the cilium as illustrated in Figure 25. A contractile impulse is believed to travel from the body of the cell outward along these filaments, causing progressive contraction. It has been suggested that the impulse might be transmitted along the filaments on one side before the other side. This would cause rapid movement of the cilium in one direction followed by slow relaxation in the opposite direction. Another type of cilia contraction, a corkscrew movement, which is utilized for forward propulsion of many unicellular animals, might be achieved by slight delay in the spread of the contractile process to the adjacent pairs of filaments. That is, one filament would begin to contract, then the adjacent one, then the next, the next, and so forth, the contraction spreading

outward along the cilium and around its axis at the same time.

REFERENCES

Edelman, I. S.: Transport through biological membranes. *Ann. Rev. Physiol.* 23:37, 1961.

Ernster, L., and Lindberg, O.: Animal mitochondria. *Ann. Rev. Physiol.* 20:13, 1958.

Harris, R. J. C., editor: *The Relationship Between Nucleus and Cytoplasm. Exp. Cell. Res.* Supplement 6, New York, Academic Press, 1959.

Lehninger, A. L.: Water uptake and extrusion by mitochondria in relation to oxidative phosphorylation. *Physiol. Rev.* 42:467, 1962.

Siekevitz, P.: Protoplasm: endoplasmic reticulum and microsomes and their properties. *Ann. Rev. Physiol.* 25:15, 1963.

Swanson, C. P.: *The Cell.* Englewood Cliffs, Prentice-Hall, Inc., 1960.

Wilbrandt, W.: Transport through biological membranes. *Ann. Rev. Physiol.* 25:601, 1963.

CONTROL OF CELL FUNCTION; AND CELL REPRODUCTION

4

One of the most important phases of cell function is the mechanisms by which the different cellular activities are regulated. These include regulation of the overall rate of cellular metabolism, the rate of synthesis of each substance produced in the cell, the rate of cell growth, the rate of cell division, and control of all other chemical processes that take place in the cell. These problems have become so important to the cell physiologist in recent years that at least one half of all research in cell physiology is now directed toward elucidating these control mechanisms.

Essentially all control mechanisms depend on *feedback*, which means that for each function performed by the cell some type of signal is fed back to the part of the cell performing this function, thereby controlling its rate of operation. For instance, let us assume that substance *X* is formed in the cytoplasm of the cell. If substance *X* should continue to be formed indefinitely at the most rapid rate that the cell could synthesize it, eventually its concentration might become so great that it would actually kill the cell. However, when the concentration approaches a definite value, this substance causes some type of feedback to the mechanisms of the cell that cause its production, thereby limiting further production and keeping its concentration in step with the concentrations of other substances in the cell. Since each cell may be synthesizing hundreds of different com-

pounds, it is evident that hundreds of separate feedback "loops" must exist in the cell, all of which act together to keep the different parts of the cell functioning in concert with each other.

CONTROL OF PROTEIN SYNTHESIS

It was pointed out in Chapters 2 and 3 that most functions of the cell are controlled either directly or indirectly by proteins, because the proteins act as enzymes which catalyze the different cellular reactions and also because the proteins comprise many of the major structural elements in the different physical parts of the cell. Therefore, control of protein synthesis in the cell indirectly controls the function of the entire cell, which means that the basic core of all cellular control systems is the control of protein synthesis itself.

Genetic Control of Protein Synthesis

The steps in protein synthesis were discussed in the preceding chapter, and it was pointed out that the type of protein that is produced by each ribosome on the endoplasmic reticulum is determined by the nucleotide composition of the ribonucleic acid (RNA) that is present in the ribosome. In turn, the nucleotide composition is determined by the genes in the nucleus in the following manner:

Chemical nature of the gene. The genes of the nucleus are very large molecules of *deoxyribonucleic acids* which are similar to the ribonucleic acids of the cytoplasm. Each deoxyribonucleic acid (DNA) is a combination of four different substances, *phosphoric acid, deoxyribose* (a sugar), and *purine* and *pyrimidine* bases. Two purine bases, *adenine* and *guanine,* and three pyrimidine bases, *cytosine, thymine,* and *methylcytosine,* are found in DNA. Large numbers of phosphoric acid and deoxyribose radicals are believed to be arranged in a helical manner around purine and pyrimidine bases as shown in Figure 26. The five different bases occur in a definite "pattern" along the axis of the DNA molecule, and this pattern in turn constitutes a "key" to the function of the gene. That is, a pattern adenine-cytosine-guanine-thymine would give a gene with a different function from that of a nucleic acid with a pattern guanine-cytosine-adenine-thymine. The molecular weight of the DNA molecule ranges from several hundred thousand up to many millions, which means that thousands of purine and pyrimidine bases may be required in the pattern of a single gene.

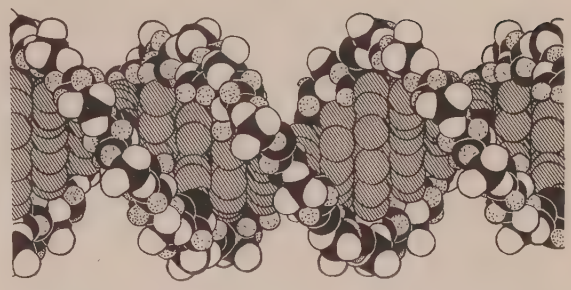


Figure 26. Suggested structure of a deoxyribonucleic acid molecule. The central shaded areas are the purine and pyramidine bases that constitute the "key" of the gene.

Arrangement of genes in the chromosomes. Each cell nucleus in the human being contains 46 chromosomes arranged in 23 pairs, and distributed within the structure of each chromosome are hundreds of different genes. When a gene of a certain type appears on one chromosome, a gene of the same or very similar type appears on the opposite chromosome of the chromosomal pair. Thus, the genes, like the chromosomes, also exist in pairs. The nucleus of the human cell has been estimated to contain a total of about 30,000 gene pairs.

The precise manner in which the genes are attached to the chromosome is still a matter of speculation. However, the most prevalent view is that represented in Figure 27. The solid block in this figure represents a structural protein which forms the "backbone" of the chromosome, and the helices attached to this structural protein are the DNA molecules, which are the genes. It is interesting that the exact function of each gene depends not only on its own chemical structure but also on its position in the chromosome.

Production of RNA by the genes. The ribonucleic acid of the ribosomes in the cytoplasm is

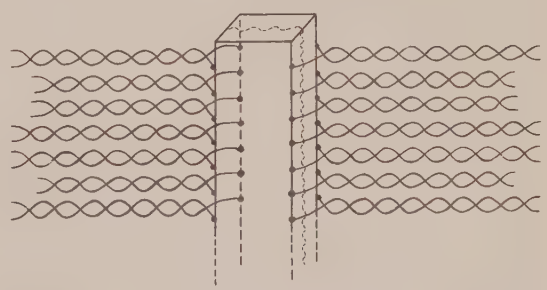


Figure 27. A suggested arrangement of the DNA molecules (the genes) on the chromosome, showing spiral DNA molecules attached by one of their ends to a structural protein that forms the backbone of the chromosome. (Redrawn from Giese: *Cell Physiology.* 2nd Ed.)

synthesized in the nucleus in direct apposition to the genes themselves. It is presumed that the genetic patterns of the DNA molecules, or mirror images of the patterns, are reproduced in the RNA molecules. In other words, the gene acts as a "template" or mold for the formation of the RNA, determining the sequence of purine and pyrimidine bases that is laid down in the ribonucleic acid. Because each gene can transmit its own pattern to a separate RNA molecule, theoretically there could be as many different types of RNA molecules as there are types of genes.

Once the RNA is formed in the nucleus, it then passes in some unknown manner into the cytoplasm to become assimilated into ribosomes and attached to the endoplasmic reticulum. Some cell physiologists believe that RNA combines with proteins in the nucleolus to form the ribosomes. Then these supposedly pass through the nuclear membrane and disperse throughout the cytoplasm. The mechanism for transfer of the ribosomes to the cytoplasm is believed to be the following: A small amount of nucleoplasm protrudes outward through the nuclear membrane, forming a pocket, and this in turn pinches off and floats freely in the cytoplasm. After a short time, the membrane around this vesicle disappears, and the contained substances disperse freely in the cytoplasm. This mechanism is similar to pinocytosis, which was described in the previous chapter as a method for transport of large molecular weight substances through the cell membrane. Other physiologists believe that the large ribonucleoprotein molecules pass directly through the large pores in the nuclear membrane.

Recapitulation of the Scheme for Control of Protein Synthesis

The basic controlling element of protein synthesis is the gene located in the chromosome of the nucleus. The gene in turn acts as a template for the formation of RNA as illustrated in Figure 28. The RNA in turn is extruded from the nucleus into the cytoplasm where it eventually attaches to the endoplasmic reticulum in the form of ribosome particles. The ribosomes then act as templates to determine the sequence of amino acids in protein molecules which are synthesized in apposition to the ribosomes as was described in the previous chapter. Thus, since the RNA molecules are patterned after the gene and the protein molecules are patterned after the RNA, it is actually the DNA molecules of the chromosomes (the genes) that control the types of protein formed in the cell.

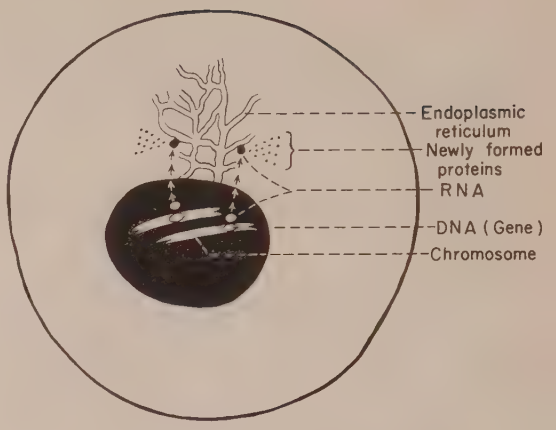


Figure 28. Control of protein synthesis.

Since there are perhaps 30,000 different gene pairs in the nucleus, the cell potentially could produce as many as 30,000 different types of protein molecules. Actually, this probably does not occur in any single cell because large numbers of the genes appear to be either totally or almost totally inactive in each cell even though the same genes might be quite active in other cells. The reasons for this are discussed at subsequent points in this chapter.

Control of the Rate of Protein Synthesis

Since all parts of the cell must function in harmony with all other parts, it is essential that the *rate* of synthesis of the different cellular elements be controlled very exactly and kept in step with each other. For this reason, protein formation does not continue unabated all the time; instead, it increases and decreases in accord with the demands of the cell. Indeed, experiments have shown that if a cell which normally produces many different types of proteins becomes unable to produce any one of these, the production of all the rest sometimes stops also to keep the *ratios* of the different types of protein from becoming abnormal. For instance, let us assume that a given amino acid is essential to the formation of a certain type of protein but that most of the other proteins of the cell can be produced without this amino acid. Experiments have shown that deficiency of this particular amino acid will stop production not only of the protein that requires it but will also stop production of all the other proteins as well. In other words, *feedback control systems* keep the production of the different types of proteins in step with each other. Let us now see how these feedback systems operate.

Enzyme inhibition. One of the most common types of feedback is called *enzyme inhibition.* Figure 29 illustrates this as follows: a gene controlling the formation of RNA, RNA controlling the formation of an enzyme, and the enzyme controlling a chemical process of the cell that determines how much of a certain chemical will be formed. If the chemical is formed in excess, it in turn acts directly on the enzyme to inhibit its function as shown by the dashed arrow number 1 in the figure. Obviously, the rate of production of the chemical then decreases, thus preventing an overabundance of the substance. An example of enzyme inhibition is the following: In some animal cells isoleucine is synthesized from threonine. However, once a small quantity of threonine has been converted to isoleucine, the enzyme system for causing this conversion becomes completely inhibited. Yet, if the enzymes are placed in an isoleucine-free medium, the synthetic process begins again immediately. Thus, isoleucine acts as a feedback product to inhibit directly the enzymes for conversion of threonine into isoleucine. Without this inhibition, essentially all of the threonine in the cell would be converted to isoleucine, and the cell would have no remaining threonine for other uses, which would be disastrous because threonine is one of the amino acids that is essential for formation of almost all cell proteins. Therefore, it is important that this feedback mechanism does limit the conversion of threonine to isoleucine and thereby maintain a relatively exact ratio between these two amino acids in the cell.

RNA and DNA repression. Figure 29 also shows two other feedback loops besides enzyme inhibition, represented by arrows 2 and 3, which show *inhibition of gene function* and *of RNA function* by chemical products produced in the cell. For instance, when excess quantities of glucose are present in the cells and this glucose becomes split into its various degradation products, certain ones of the degradation products act on some of the genes or their corresponding RNA's to decrease their rates of activity. This decreases the rates of formation of some of the protein enzymes that control utilization of glucose and other nutrients in the cell, such enzymes as beta-galactosidase, inositol-dehydrogenase, glycerol-dehydrogenase, histidase, tryptophanase, and amino acid deaminases.

Thus far, it has not been possible to tell in the case of most of these "repression mechanisms" whether the repression occurs at the genetic level or at the RNA level. Nevertheless, it is by such feedback mechanisms that the quantities of enzymes in the cell are controlled to meet the needs of the cell. Thus, if excess glucose is available, then the enzymes for utilizing glucose and other nutrients are reduced. On the other hand, if too small a quantity of glucose is available, these enzymes are increased so that the small amount of glucose that is available can be utilized more easily and so that other substances such as proteins might also be utilized for energy in place of glucose. This is an automatic mechanism of the cell, which allows continued normal energy metabolism in the cell despite vast changes in the availability of one of the cellular nutrients.

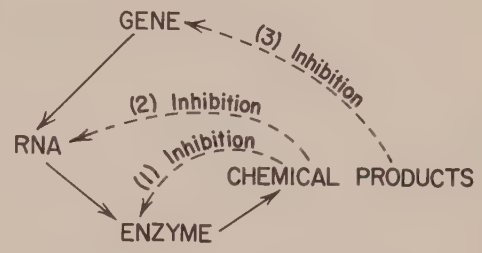


Figure 29. Feedback control systems in the cell that control *rates* of synthesis of the different cellular proteins. Arrow 1 represents enzyme inhibition, while arrows 2 and 3 represent repression of genetic and RNA activity.

To summarize, each chemical product, by enzyme inhibition, by RNA repression, or by DNA repression, limits its own concentration in the cell, and whenever its concentration falls too low, opposite effects occur, so that increased quantities of the substance are produced. Each feedback mechanism for each chemical substance in the cell has its own critical level of sensitivity so that some chemical products are produced in great quantities in the cell before they become limited while others are produced only in small quantities.

Control of protein synthesis is so important that some of the genes in the nucleus actually control the rates of activity of other genes. Also, in the developing embryo, as we will discuss later in the chapter, local hormones called *evocators* often pass from one cell of the embryo to other cells to control their rates of protein synthesis. And, in the adult human body, we know that hormones elaborated in certain parts of the body flow in the body fluids to other parts of the body where they control intracellular protein synthesis, thereby controlling some of the functions of the cells. Thus, control of protein synthesis not only is a means by which intracellular function is regulated, but is also a means by which certain functions of the different parts of the animal body are kept in step with each other.

CONTROL OF ENERGY METABOLISM IN THE CELL

The term *metabolism* means all the chemical reactions that occur in the cell, and the term *energy metabolism* means the chemical reactions that supply energy to the cell for performing its functions.

We noted in the previous chapter that tremendous amounts of energy are used in different cells for (a) synthesis of new chemicals, (b) membrane transport, and (c) muscle contraction.

The overall metabolic rate (rate of energy utilization) of the cell is determined by the aggregate amount of energy expended in all these processes. For instance, when glandular cells are stimulated either by nerve impulses or by hormones to form their secretions, the chemical reactions in the glandular cells proceed at vastly increased rates, and the synthesized products are extruded from the cell. Obviously, large amounts of extra energy are required to promote these chemical reactions. This energy is derived from adenosine triphosphate, as was pointed out in the previous chapter.

Another type of activity that requires tremendous amounts of energy is membrane transport during the transmission of nerve impulses. As each impulse travels along a nerve fiber a large amount of sodium diffuses into the fiber; after the impulse has passed, this sodium must be transported back out of the fiber into the surrounding extracellular fluid. The presence of the excess sodium in the cell automatically stimulates the "sodium pump" to cause its transport back out of the cell. Simultaneously, large amounts of energy are derived from adenosine triphosphate to make the necessary reactions occur.

Likewise the rate of energy expended by muscle fibers increases tremendously when they contract, sometimes as much as 100-fold for a few seconds at a time. And, here again, the energy comes from adenosine triphosphate.

Control of nutrient usage in response to the metabolic rate. If the metabolism is to change markedly from time to time in the cell, it is important that the amount of energy released from the nutrients changes accordingly. Figure 30 illustrates the means by which glucose utilization in the cell is controlled. But before explaining this diagram, let us recall that all the different energy-using processes of the cell derive their energy from *adenosine triphosphate* which is converted to *adenosine diphosphate*. Adenosine diphosphate is therefore the key to the control of energy release from the nutrients, because unless it is present in the cell none of the energy-giving reactions involving the nutrients can occur. In the case of glucose, adenosine diphosphate must interact with some of the intermediate compounds in the sequence of glucose breakdown before the reactions will proceed further. Then, when adenosine diphosphate does interact, it is later released in the form of adenosine triphosphate which has gained energy removed from the glucose molecule as it is split into its smaller particles.

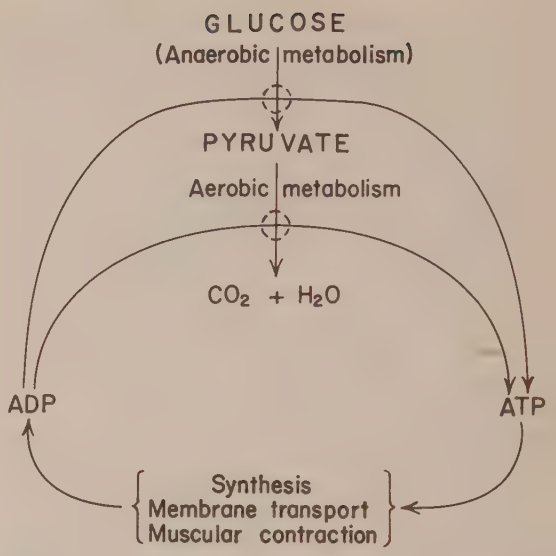


Figure 30. The basic mechanism by which cellular use of nutrients is controlled.

Therefore, in essence, the rate of use of glucose is controlled by the rate of metabolism in the following way: The greater the rate of metabolism the greater the amount of ADP formed. In turn, the greater the amount of ADP formed, the greater becomes the rate of glucose breakdown and also the rate of reformation of ATP. Thus, the cell automatically removes ADP whenever it is formed and thereby maintains an excess of ATP, keeping this always available to supply energy wherever in the cell it is needed. The breakdown of lipids and proteins for delivery of energy to the cell is also controlled in almost identical ways.

CELL GROWTH AND ITS CONTROL

The cell grows from within, each part being continually synthesized and increasing in quantity. For reasons that are not yet understood, when the cell reaches a certain size it automatically divides, forming two new cells, a process called *mitosis* that is described later in the chap-

ter. Cell growth proceeds unabated as long as appropriate nutrients are available and as long as the cell is not inhibited by some external factor. Control of cell growth is exceedingly important to the human body because this determines the sizes of the cells in the organs, the numbers of cells in each organ, and indirectly even the functions of the organs.

Regulation of Cell Growth

In the adult human being, the *ratios* of numbers of different cells in different parts of the body are almost constant. For instance, if for any reason too many cells of one type develop, then, these will stop growing and dividing until some of the cells of this type have died and new ones are needed. On the other hand, if for some reason there should become too few of one type of cell, growth and division of this type of cell will proceed unabated until appropriate numbers of them are again available. For instance, ⅞ of the liver of the human being can be removed surgically, and the cells in the remaining ⅛ will grow and divide until the liver mass is almost back to normal. This same effect occurs for almost any of the glandular cells, for cells in the bone marrow, in the subcutaneous tissues, in the intestinal epithelium, and in almost any other tissue—indeed for essentially all cells except a few very highly differentiated types such as the nerve cells. This principle applies also to the growth and reproduction of cells in the embryo human being.

We know almost nothing about the regulation of cellular growth and about the various factors that maintain the proper ratios of different cells. Studies in embryos indicate that some growing cells affect adjacent cells by secretion of local hormones that diffuse to the adjacent cells and control intracellular protein synthesis mechanisms.

It is assumed that similar hormones might be present even in the adult human being, though these have not been discovered. We also do not know why cells stop growing and dividing when their numbers reach a certain level. Perhaps each type of cell secretes some substance that has a direct feedback effect on that same type of cell to stop its growth, though such substances have not yet been found. We do know that cells of any type removed from the body and grown in tissue culture can continue to grow and divide indefinitely if the medium in which they grow is continually replenished. Yet, the cells will stop growing when secretions from the cells are allowed to collect in the medium.

Cancer

Cancer can occur in any tissue of the body. It results from a change that allows the cancerous cells to disrespect normal growth limits, no longer obeying the feedback control systems that normally stop cellular growth after a certain number of such cells have developed. As pointed out above, even normal cells when removed from the body and grown in tissue culture can grow and proliferate indefinitely if the growth medium is continually changed. Therefore, in tissue culture there seems to be no difference between normal tissue cells and cancer cells, but in the body it is entirely different, for here the normal tissue cells are subject to feedback limits while cancer cells are not.

What is the difference between the cancer cell and the normal tissue cell that allows the cancer cell to grow and divide unabated? The answer to this question is not known, but researchers have found the genetic structure of cancer cells to be different from that of normal cells. This has led to the idea that cancer probably results from mutation of some of the genes during one of the cellular division processes, these mutations eliminating the feedback mechanisms that normally limit growth of the cell. Once even a single such cell is formed, obviously it will grow and proliferate in the human body in exactly the same way that it grows and proliferates in tissue culture, its number increasing exponentially.

Once a cancer begins it competes for nutrients with the animal's normal tissues, and because cancer cells continue to proliferate exponentially, their number multiplying day by day, one can readily see that the cancerous cells will soon demand and utilize essentially all the nutrition that can be made available by the body. As a result, the animal's own tissues gradually suffer nutritive death.

CELL REPRODUCTION

Mitosis

Except in rare instances, cell division occurs by the process known as *mitosis,* which leads to the formation of two daughter cells having characteristics identical with those of the parent cell. During mitosis, both the cytoplasm and the nucleus divide approximately equally between the two daughter cells. The stages of mitosis are so well known to most students from their elementary courses in biology that they are pre-

sented here only briefly. However, some aspects of mitosis that are of particular importance from a functional point of view are discussed in more detail.

Factors that determine when mitosis will take place. Mitosis seems to be a natural result of growth, for whenever cells grow to a certain size, they automatically undergo the process of mitosis. Precisely what determines the critical size at which mitosis will occur is not known. One of the theories suggests that this size is determined by the *ratio of the cell's volume to its surface area,* which may be explained as follows: The nutritive needs of a cell are determined by its volume, while the amount of nutrients available are determined by the surface area through which the nutrients can be transported. Yet, the volume increases in proportion to the cube of the cell diameter, while the surface area increases in proportion to the square of the diameter. Therefore, the larger the diameter of the cell, the greater becomes the ratio of the needed nutrients to the available nutrients. It could be that this relative decrease in available nutrients might be the stimulus for cell division. One type of experiment that indicates this to be true is the following: A cell that has grown almost large enough to divide can be prevented from doing so by surgical removal of a large segment of the cytoplasm. The cell will then grow once again and still will not divide until it reaches the usual critical diameter.

Rates of mitosis. Growth and reproduction of cells is extremely rapid in the embryo, the mitotic cycle often recurring once every few hours, but this slows progressively as the embryo grows larger. By the time the adult human being has been formed, the rate of mitosis is so slow that only rare mitotic cells are found in most tissues of the body. Neurons, for instance, have lost entirely their capability to undergo mitosis; also, muscle cells, tendon cells, and cells in specialized organs such as the kidneys all have mitotic cycles that can be measured in months or years. Yet, mitosis continues at rapid rates in a few tissues such as in the blood-forming tissues, the germinal layers of the skin, and the epithelium of the gut.

When the cells from any tissue of the body are removed and grown in tissue culture, they grow and divide unabated. Therefore, in essence, the factor that determines the rate of mitosis of the cells in the body are the growth-limiting feedback controls that were discussed previously.

Replication of the genes. Before mitosis will take place, all the genes are duplicated so that each daughter cell will contain a complete compliment of genes. At some time during *interphase* (the period between mitoses) the DNA molecules that constitute the genes are replicated. That is, a single, exactly duplicated molecule of DNA is formed for each gene, using the original gene as a model to determine its structure. Only a single new gene is formed for each of the original genes, and despite the complexity of the DNA molecule, this replication process rarely makes a mistake. During the process of mitosis each pair of duplicates becomes evenly divided between the two daughter cells.

Replication of the chromosomes. Not only is each gene replicated, but the entire structure of each chromosome is also duplicated, with the different genes attached to the chromosome in the same order as in the original chromosome. The mechanism of this replication is yet unknown.

The stages of mitosis. Figure 31 illustrates the physical changes in the cell during mitosis. The first effect is division of the *cell center* which lies

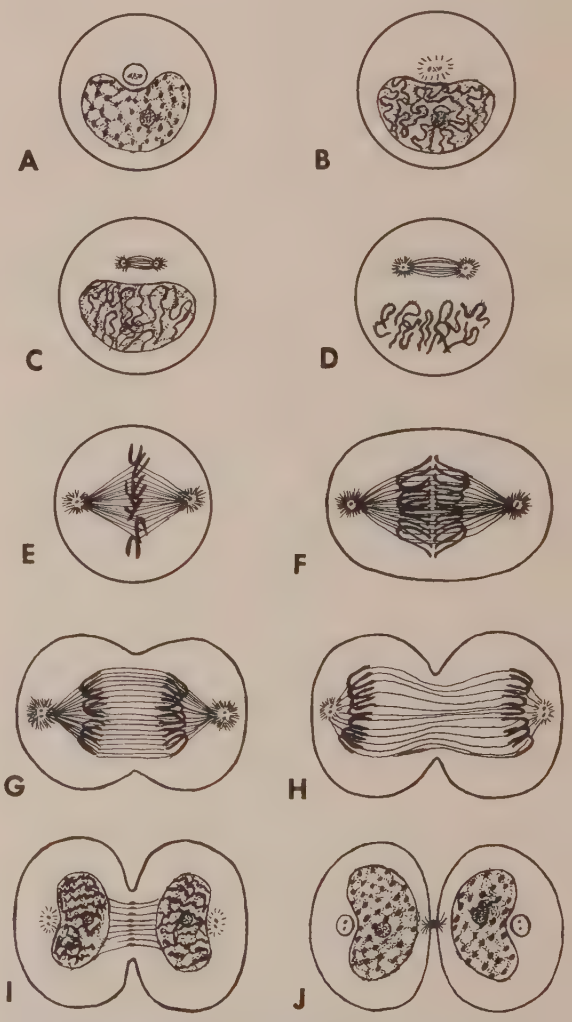


Figure 31. The different stages in mitosis. (From Bloom and Fawcett: *A Textbook of Histology.* 8th Ed.)

in the cytoplasm immediately outside the nucleus. As illustrated in Figures 31A, B, C, and D, the two new centers pull apart, thereby initiating the process of mitosis. This begins with the development of a mitotic spindle as shown in Figures 31B and C; this stage of mitosis is called *prophase.* During prophase, the small chromatin particles of the nucleus line up in threads to form well delineated chromosomes as illustrated in Figure 31C. Some physical structure probably exists even during interphase to maintain integrity of the chromosomes, though this cannot be seen even with the electron microscope.

Figure 31D illustrates *prometaphase* during which the nuclear membrane disappears and the chromosomes begin moving into the mitotic spindle. Figure 31E represents *metaphase* during which the chromosomes line up on the equitorial plane of the mitotic spindle. Furthermore, by this time it becomes evident that each chromosome consists of two duplicates known as the *chromatids.* One of the chromatids faces one of the cell centers and the other faces the opposite cell center.

During *anaphase,* illustrated in Figures 31F, G, and H, the chromatids, which are now known as the *daughter chromosomes,* move along the mitotic spindle toward the separate poles. At the same time, the *centriole* in each cell center divides to form two new centrioles, which will form the basis for the respective cell centers of the daughter cells during the next mitosis.

During *telophase* (Figure 31I) each group of daughter chromosomes fuses into a mass around which a new nuclear membrane develops. Simultaneously, the cell begins to constrict in the middle until it breaks to form two new daughter cells, as shown in Figure 31J.

REGULATION OF THE STEPS IN MITOSIS. The causes of the different steps in mitosis are completely unknown. Many cell physiologists have speculated that electrical gradients develop between the cell centers, this causing the chromosomes to move as they do. However, another theory is that special tubular systems develop that result in the movements. Unfortunately, even after study with the electron microscope we still know far too little of the basic structures involved in mitosis to give even a hint of the factors that cause movement of the chromosomes.

Yet, the purpose of mitosis is quite evident: to separate the replicated chromosomes from each other so that each daughter cell will receive a full complement of chromosomes. Therefore, whatever the factors that control mitosis, they must, first, orient the chromosomes in such a manner that the two chromatids can be pulled apart. Once pulled apart, two new nuclei must be reformed and the cell split as described above for the process of mitosis.

CELL DIFFERENTIATION

A special problem in cell growth and cell division is *cell differentiation,* which means changes in physical and functional properties of cells as they proliferate in the embryo to form the different bodily structures. Figure 32 illustrates various stages of cleavage of the fertilized mammalian ovum, the darkened cells representing those that eventually will form the embryo and the lighter cells those that eventually will form the fetal membranes. Note the difference in sizes and other characteristics of the cells as the cell division proceeds during the first few days after fertilization. Obviously, still many more changes in characteristics of the cells will take place before the final postnatal human being has been formed. However, our problem is not to describe the stages of differentiation but simply to discuss the theories of how the cells progressively change their characteristics and form all the different organs and tissues of the body during gestation.

The earliest and simplest theory for explaining differentiation was simply that the genetic composition of the nucleus might undergo changes during successive generations of cells in such a way that one daughter cell would develop one set of characteristics while another daughter cell would develop entirely different characteristics. This theory probably still explains some types of differentiation because highly differentiated cells cannot be reverted all the way back to the original primordial state. Yet, most cells can be reverted at least two or three steps backward in the process of differentiation, which indicates that other factors besides simple changes in genetic potency probably also play a role in differentiation and might even play the dominant role.

The cytoplasm also probably plays a role in differentiation, because cells without nuclei can divide and even differentiate to some extent for



Figure 32. Differentiation of cells in the early embryo during gestation: (A) two blastomeres, (B) six blastomeres, (C) a hemisected early blastocyst, (D) a later blastocyst stage showing the early fetus at the top. (Modified from Arey: *Developmental Anatomy.* 6th Ed.)

a few stages before the cell dies. However, the role of cytoplasm in differentiation is probably controlled by the nucleus.

Embryological experiments show that certain cells in an embryo control the differentiation of adjacent cells. For instance, the *primordial chorda-mesoderm* is called the *primary organizer* of the embryo because it forms a focus around which the rest of the embryo develops. It differentiates into a mesodermal axis containing segmentally arranged *somites*, and as a result of complex *inductions* in the surrounding tissues causes formation of essentially all the organs of the body.

Another instance of induction occurs when the developing eye vesicles come into contact with the ectoderm of the head and cause it to thicken into a lens plate which folds inward to form the lens of the eye. It is possible that the entire embryo develops as a result of complex inductions, one part of the body affecting another part, and this part affecting still other parts.

Thus, our understanding of cell differentiation is still very hazy. We know many different control mechanisms by which differentiation *could* occur. Yet, the overall basic controlling factors in cell differentiation have not been discovered; when learned, they will make a tremendous difference in our understanding of bodily function.

REFERENCES

Beadle, G. W.: Physiological aspects of genetics. *Ann. Rev. Physiol.* 22:45, 1960.

Ciba Foundation, Symposium Volume: *Biochemistry of Human Genetics.* Boston, Little, Brown and Company, 1960.

Glaser, L.: Biosynthesis of deoxysugars. *Physiol. Rev.* 43:215, 1963.

Moyed, H. S., and Umbarger, H. E.: Regulation of biosynthetic pathways. *Physiol. Rev.* 42:444, 1962.

Prescott, D. M.: Nuclear function and nuclear-cytoplasmic interactions. *Ann. Rev. Physiol.* 22:17, 1960.

Sohval, A. R.: Chromosomes and sex chromatin in normal and anomalous sexual development. *Physiol. Rev.* 43:306, 1963.

Stern, H.: Function and reproduction of chromosomes. *Physiol. Rev.* 42:271, 1962.

Zamenhof, S.: *The Chemistry of Heredity.* Springfield, Charles C Thomas, 1959.

THE BODY FLUIDS AND THE URINARY SYSTEM



THE FLUIDS OF THE BODY

5

The fluid inside the cells, called the *intracellular fluid,* provides the medium in which the cell's metabolic reactions take place, while the fluid outside the cells, called the *extracellular fluid,* provides the medium in which the cells live. If the concentrations of the extracellular fluid constituents remain at appropriate levels, the cells will continue to live indefinitely. Therefore, one of the most important functions of the body is to maintain constant conditions in these fluids, a function called *homeostasis,* the importance of which was discussed in Chapter 1. In the present chapter, we will characterize both the intra- and extracellular fluids and will begin our discussion of the mechanisms by which constancy is maintained in the fluid environment of the cells.

BODY WATER, INTRACELLULAR FLUID, AND EXTRACELLULAR FLUID

The total water in the average 70 kilogram man is about 40 liters or 57 per cent of his total body weight. This water is divided between two major compartments, the *intracellular compartment,* which is all the fluid inside the cells, and the *extracellular compartment,* which is all the fluid outside the cells. On the average, about 62 per cent of the total body water is inside the cells and about 38 per cent is outside. Thus, the average

70 kilogram man has about 25 liters of intracellular fluid and about 15 liters of extracellular fluid, as shown in Figure 33.

Blood. About 97 per cent of the blood is fluid. However, this fluid, like the total body water, comprises intracellular and extracellular fluids. The *plasma* portion of the blood, which is the clear fluid between the cells, is typical extracellular fluid except that it has a higher concentration of protein than do the extracellular fluids elsewhere in the body, which is very important for keeping the plasma inside the circulatory system, as is discussed in the following chapter. The *cells* are of two types, *red blood cells* and *white blood cells.* The number of white blood cells is only one five-hundredth the number of red blood cells so that, for practical purposes, essentially all of the intracellular fluid in blood is contained in the red blood cells. The normal total volume of plasma is 3 liters and of red cell fluid 2 liters.

Interstitial fluid. The interstitial fluid is that portion of the extracellular fluid that lies outside the capillaries and between the cells. The volume of the interstitial fluid, therefore, is equal to the total extracellular fluid volume minus the plasma volume, or $15 - 3 = 12$ liters of interstitial fluid volume in the normal, average adult man. Interstitial fluid is usually considered to include such special fluids as those in the cerebrospinal fluid system, the fluid in the eyes, in the minute intrapleural spaces, in the peritoneal cavity, in the pericardial cavity, and in the joint spaces.

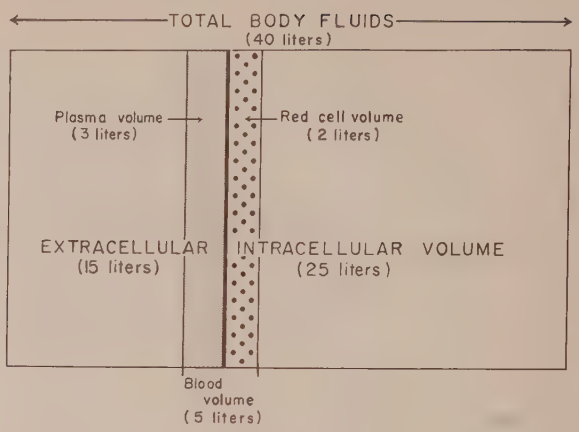


Figure 33. Diagrammatic representation of the body fluids, showing the extracellular fluid volume, blood volume, and total body fluids.

Constituents of the Intracellular Fluid

Intracellular fluid contains literally hundreds of organic compounds that are responsible for the intracellular chemical reactions discussed in the past few chapters. In addition to these compounds, the intracellular fluid contains, as illustrated in Figure 34, large quantities of the electrolytes *potassium, magnesium,* and *phosphate* plus small quantities of *sodium, chloride, bicarbonate,* and *sulphate.* Some of these electrolytes are extremely important as integral parts of the chemical reactions occurring in the cell. It was pointed out in Chapter 3 that the cell membrane has special capabilities for actively transporting many electrolytes between the intracellular and extracellular fluid. Most important of these are the *sodium pump which transports sodium out of the cell* and the mechanisms for *transport of potassium and phosphate into the cell.*

Intracellular fluid also contains an assortment of nutrients to be used in the cellular metabolic processes. Included in these is a small amount of *glucose* which is mostly converted to intermediate metabolic products for further use in the cell or is stored in the cell in a polymerized form as insoluble *glycogen.* Most cells also contain small quantities of *lipids,* including *neutral fat, phospholipids,* and *cholesterol,* and the cells of fat tissue may contain as much as 95 per cent of their total mass as fat. However, most of the lipids are insoluble in water and, therefore, are present in the cell in a colloidal or deposited form rather than dissolved in the fluids. Another important nutrient in the cell is the *amino acids,* which are probably present in greater concentration than in the extracellular fluid because of active transport through the cell membrane.

Intracellular fluid also contains *oxygen* in a reasonably high concentration. As the oxygen is metabolized with the nutrients to give energy, *carbon dioxide* is liberated into the intracellular fluid, and it then diffuses outward through the cellular membrane into the extracellular fluid.

Constituents of the Extracellular Fluid

Most striking of all the differences between the extracellular and intracellular fluids are the concentrations of the electrolytes. The *sodium, calcium,* and *chloride* concentrations are many times as high in the extracellular fluid as in the intracellular fluid; in contrast, the potassium, magnesium, and phosphate concentrations are many times as high intracellularly as extracellularly. These differences are responsible for electrical potentials that develop across the cell membrane and also for the degree of permeability of the membrane. We shall see in Chapter 20 that it is mainly the concentration gradient of potassium

between the two sides of the membrane, 141 milliequivalents per liter intracellularly and 5 milliequivalents per liter extracellularly, that causes electrical potentials to develop across the membranes of nerve fibers, muscle fibers, and probably also of all other cells of the body. Without these membrane potentials, it would be impossible for nerve fibers to conduct impulses or muscle fibers to contract.

The calcium in the extracellular fluid plays a role in determining the degree of permeability of the membrane. A high concentration of calcium decreases the permeability while a low concentration increases the permeability. Furthermore, whenever a cell membrane ruptures, if calcium is present in large enough quantities, a new membrane will be formed immediately, which is called the "surface precipitation reaction"; this was described in Chapter 2.

The nutrients in the extracellular fluid, like those in the intracellular fluid, include glucose, amino acids, and lipids. However, the concentrations of these in the extracellular fluid are quite different from the concentrations inside the cell. For instance, the extracellular glucose concentration is very high in comparison with that in the intracellular fluid, as shown in Figure 34. However, if we should consider all the *products* of glucose inside the cell—hexose monophosphate, glycogen, and so forth—we would find that the cell actually concentrates glucose in these other forms in very great quantities to be used for energy when needed. Likewise, the concentration of amino acids in the extracellular fluid is low compared to the amount of amino acids inside the cell. Here again few free amino acids are present inside the cell; instead, they are conjugated into simple proteins that can be hydrolized rapidly back into amino acids. Also, lipids are usually present in the cell in far greater quantities than in the extracellular fluid, usually deposited in membranes or in inclusions within the cytoplasm.

Finally, oxygen is present in the extracellular fluid in slightly higher concentration than in the intracellular fluid. It is this concentration difference that causes oxygen to diffuse continually from the extracellular fluid into the cell. And *carbon dioxide* is present in intracellular fluid in slightly greater concentration than in extracellular fluid, which allows carbon dioxide to diffuse continually out of the cell as it is formed by the combustion of nutrients with oxygen.

To summarize, the constituents of extracellular fluid have two purposes: first, to maintain membrane potentials and proper membrane permeability so that very important membrane functions can take place, and, second, to provide a system for transport of nutrients and other substances from one part of the body to another.

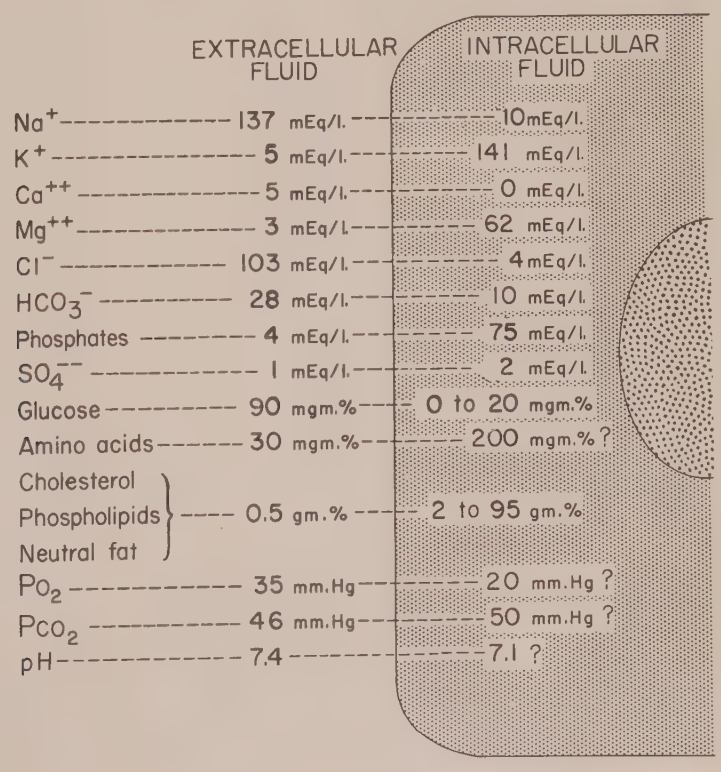


Figure 34. Chemical compositions of extracellular fluids.

For instance, glucose, amino acids, cholesterol, phospholipids, neutral fat, carbon dioxide, urea, lactate, creatinine, sulfate, and many other substances are all transported by movement of the extracellular fluid.

MOVEMENT AND MIXING OF THE EXTRACELLULAR FLUIDS THROUGHOUT THE BODY

Fortunately, the extracellular fluids are continually being mixed. Were it not for this, the cells would continue to remove nutrients from the fluids in their immediate vicinity until none would be left, and cellular excreta would accumulate in their immediate vicinity until the excreta themselves would kill the cells. Yet, two different mechanisms provide for continual mixing and movement of the extracellular fluid throughout the body. These are the *circulation of the blood* and the phenomenon of *diffusion.* The circulatory system transports the fluid long distances in the body, while the diffusion mechanism allows rapid mixing of the fluids in the capillaries with those in the tissue spaces.

Transport of Fluid by the Circulatory System

Figure 35 illustrates the general plan of the circulation, showing continuous flow of blood around the system, through the lungs, and through the different tissues. The blood picks up oxygen and the various nutrients in the lungs and gut respectively, and then carries these to all other areas of the body. And, as the blood passes through the tissues, it picks up carbon dioxide, urea, and other excreta from the cells and transports these to the lungs and kidneys respectively where they are removed from the body. Thus, the circulatory system provides long-distance transport of the extracellular fluids, in this way keeping the fluids in the different parts of the body mixed with each other.

The rate of systemic circulation of the blood, which is called the *cardiac output,* is approximately 5 liters per minute, a volume about equal to the total blood volume. Therefore, approximately once per minute the entire blood volume makes the circuit. It is obvious from the schema in Figure 35 that each time the blood goes around the circuit, all of it goes through the lungs where oxygen is picked up and carbon dioxide is removed. Also, one fifth of the total blood flow goes through the kidneys, where cellular excreta such as urea, uric acid, creatinine, phosphate, sulfates, and so forth are removed, and where the concentrations of the different extracellular electrolytes such as sodium, chloride, calcium, potassium, and magnesium are adjusted.

Diffusion of Fluids and Solutes Between the Capillaries and the Cells

The principle of diffusion. Diffusion means the continual movement of individual molecules, ions, and other particles in all directions in a fluid. This results from the fact that each particle has a certain amount of kinetic energy; that is, it is endowed with motion. It moves until it strikes some other particle, then bounces away until it hits still another particle, whereupon it bounces again and thus continues to bounce indefinitely through the fluid. Figure 36 illustrates the movement of a single particle among many other particles in the fluid, showing that it bounces first in one direction, then another, another, and so forth, changing its direction of movement each time it collides with a water molecule or with another dissolved particle. Thus, each particle eventually finds its way into essentially all corners of the fluid space. This is the phenomenon of *diffusion,* and it applies to each single molecule or particle in the fluid whether it be dissolved or suspended. Obviously, the process of diffusion is an excellent mixing device.

Transfer of water and solutes to and from the cells by diffusion. Figure 37 illustrates a capillary in a bed of tissue cells, and the arrows in the capillary, through the capillary wall, in the tissue spaces, and through the cell walls all indicate the process of diffusion, showing that molecules, both water molecules and molecules of dissolved substances, are bouncing continually in all these different directions. Therefore, by the process of diffusion, the fluids of the capillary are mixing continually with the fluids of the interstitial spaces, and the fluids of the interstitial spaces are mixing continually with the fluids of the cells. The only factor that limits this diffusion process is the sizes of the pores at the different membranes. At the capillary membrane all particles of the plasma except the plasma proteins can diffuse into the interstitial spaces. At the cellular membrane, water, very low molecular weight substances, or fat-soluble substances such as carbon dioxide and oxygen diffuse freely into and out of cells. However, high molecular weight substances such as most of the organic compounds of the intracellular fluid and positively charged particles such as

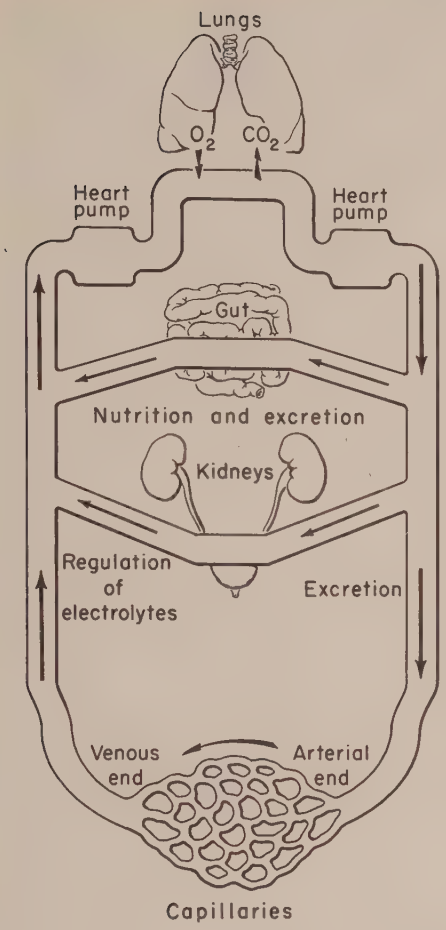


Figure 35. General schema of fluid flow in the circulation.

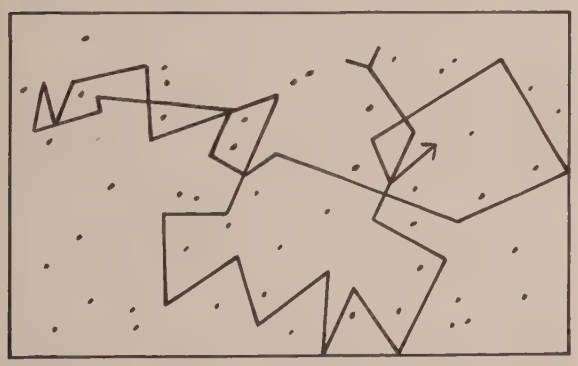


Figure 36. Diffusion of a fluid molecule during a fraction of a second.

sodium and potassium pass through the cellular pores so slowly that they do not become mixed between the intra- and extracellular fluids. Except for these limitations, water and at least certain dissolved substances mix freely and continually between the capillary, the interstitial fluid, and the fluid inside the cells.

THE "DIFFUSION GRADIENT." To illustrate the importance of diffusion for transport of nutrients to the cell, let us consider the movement of oxygen through the tissue fluids. There is no special mechanism for transport of oxygen from the capillary into the interstitial fluid or from the interstitial fluid into the cells except by this process of diffusion. The cells are continually using oxygen for combustion with glucose, fats, and proteins, which keeps the oxygen concentration inside the cells much lower than that in the blood and interstitial fluids. Yet, because of diffusion, new molecules of oxygen continually wend their way out of the capillaries toward the cells, thus replenishing the oxygen that is being used. In this way, the diffusion process moves oxygen from the area of high concentration in the capillary to the area of low concentration in the cell.

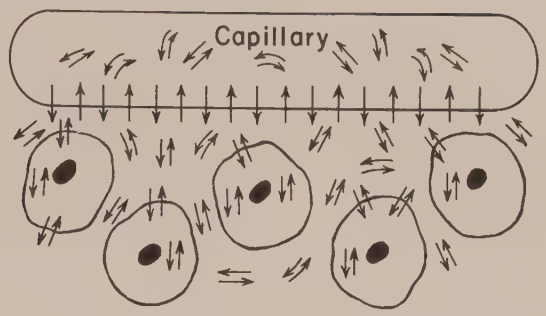


Figure 37. Diffusion of water molecules and molecules of dissolved substances in the blood, through the capillary membrane, in the tissue spaces, through the cell membrane, and inside the cell.

It is said that a *diffusion gradient* exists for oxygen between the capillaries and the cells, meaning by this that since the concentration is less in the cells than in the capillaries, more oxygen will diffuse from the capillary into the cells than in the opposite direction. In other words, the *net* movement of oxygen is along the diffusion gradient. There is also a diffusion gradient for glucose from the capillaries to the cells, because its concentration is far greater in the capillaries than in the cells. On the other hand, in the case of a substance being formed continually in the cells, such as carbon dioxide, sulfates, phenols, and so forth, diffusion gradients exist from the cells toward the capillaries, and as a consequence these substances move away from the cells rather than toward them.

Effectiveness of Extracellular Fluid Movement and Mixing

The two mechanisms for movement and mixing of extracellular fluids are so effective that almost every portion of the extracellular fluid, even that

in the minutest tissue space, becomes mixed with all the other extracellular fluid of the body at least once every 10 to 30 minutes. Because of this rapidity of mixing, the concentrations of substances in the extracellular fluid of one portion of the body are rarely more than a few per cent different from those anywhere else in the extracellular fluid. Also, a substance secreted into the extracellular fluid anywhere in the body will be distributed throughout the extracellular spaces within a few minutes. It is in this way that hormones from the thyroid gland, the adrenal gland, the sex glands, and so forth are transported to all the cells to perform their control functions, that oxygen is transported from the lungs to the cells, and that excreta from the cells are carried either to the lungs or kidneys to be removed from the body.

OSMOSIS AT THE CELL MEMBRANE

Dissolved substances that do not diffuse through the cell membrane can under certain conditions cause osmosis of water at the cell membranes, resulting in either swelling or shrinkage of the cells. Therefore, it is important to the physiologist to know the principles of osmosis.

The Principles of Osmosis

Figure 38 illustrates a cell membrane with molecules of water and dissolved substances on each side. Some of the dissolved substances have molecules that are too large to go through the pores. The molecules normally present in the intra- and extracellular fluids that can pass through the pores with complete ease are water, oxygen, and carbon dioxide. Essentially all other substances in the body fluids have considerable difficulty passing through the cell membrane. For instance, glucose molecules and the organic molecules of the intracellular fluid are all too large to diffuse through the pores. And ions fail to go through the membranes because of their electrical charges. The positively charged ions, sodium, potassium, magnesium, and calcium, do not pass through the pores because the pores themselves are lined with electropositive charges that repel the positively charged ions, as was discussed in Chapter 3. The negatively charged ions, chloride, bicarbonate, and lactate, are small enough to pass through the pores but do not do so in great quantities because, if they should, they would build up an intense electronegative charge on the opposite side of the membrane that would repel them back where they came from. In other words, unless positive ions go through the membrane to balance the electronegative charges, negative ions also cannot go through the membrane. Therefore, in essence, even though negative ions are small enough to go through the pores, in practice they do not do so.

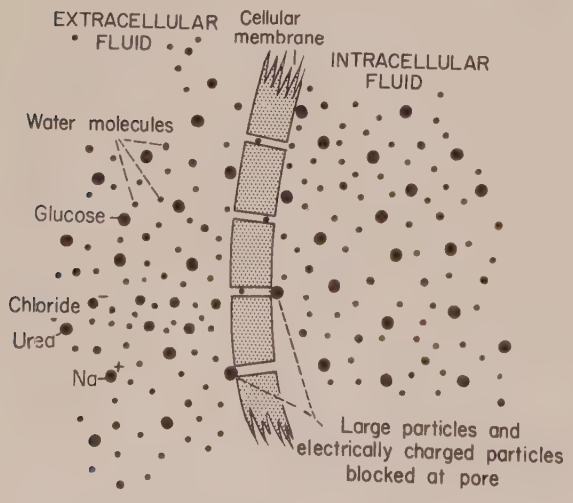


Figure 38. Relationship of the cell membrane to the molecules of the extracellular and intracellular fluid, showing passage of water molecules and the failure of other molecules to go through the pores.

Thus, we find that the only substances that pass through the cell membrane with complete ease are water molecules, oxygen, and carbon dioxide. Because oxygen and carbon dioxide are present in the body fluids in such minute concentrations, these can be ignored from an osmotic point of view.

The number of water molecules that diffuse from the extracellular fluid into the intracellular fluid each second is the number of water molecules that strike pores on the outside of the membrane and are moving in proper directions and with enough kinetic energy to carry them on through the pores to the inside of the cell. Likewise, the number of intracellular water molecules that diffuse outward is equal to the number of water molecules that similarly strike pores on the inside of the membrane. If the numbers of molecules striking the pores on the inside and outside are precisely equal, then equal amounts of water will diffuse inward and outward in a given period of time, and there will be no net volume change in either the extracellular or intracellular fluid. However, if the number of molecules striking one side of the membrane is greater than that striking the other side, more molecules of water will diffuse in one direction than the other, which is the phenomenon called *osmosis*.

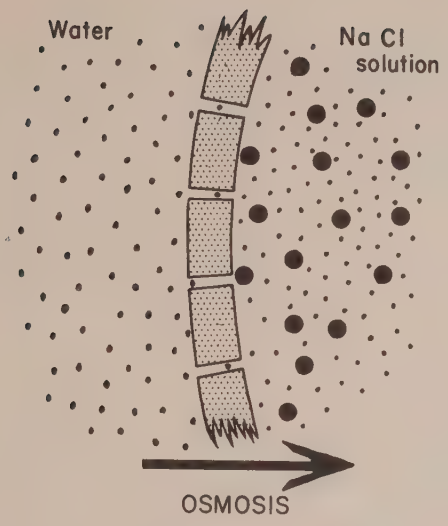


Figure 39. Osmosis occurring at a hypothetical cellular membrane when a sodium chloride solution is placed on one side of the membrane and water on the other side.

The condition that causes osmosis is a greater concentration of non-diffusible dissolved substances on one side of the membrane than on the other. For instance, let us assume that pure water is on one side of a cell membrane and sodium chloride solution is on the other side, as illustrated in Figure 39. In this case, the sodium and chloride ions fail to pass through the pores. Yet, the presence of these particles dilutes the water molecules and decreases the number of these molecules that strike the membrane in a given period of time. For this reason, considerably fewer molecules of water will diffuse out of the sodium chloride solution through the membrane than will diffuse out of the pure water in the other direction. Thus, *net* movement of water will occur from the pure water into the sodium chloride solution, which is the phenomenon of osmosis.

Osmotic Equilibrium of Extracellular and Intracellular Fluids

Except for very short periods of time, usually measured in seconds, the intracellular and extracellular fluids remain in constant osmotic equilibrium; that is, the concentrations of non-diffusible substances on the two sides of the membranes remain almost exactly equal. The reason for this is that, if ever the concentrations become unequal, osmosis of water through the cell membranes occurs so rapidly that equilibrium will be reestablished within a few seconds. This effect is illustrated in Figure 40. Figure 40A shows a cell suddenly placed in a very dilute solution (*hypotonic solution*). Within a few seconds, water passes by osmosis through the cell membrane to the inside of the cell, causing: (1) increase in the intracellular fluid volume, and thus swelling of the cell, (2) decrease in the extracellular fluid volume, (3) dilution of the dissolved substances in the intracellular fluid, and (4) concentration of the dissolved substances in the extracellular fluid. Once these two fluids have reached the same concentrations of osmotically active substances, osmosis through the membrane ceases. Thus, within a few seconds, a new state of osmotic equilibrium has been reestablished, as shown in Figure 40B.

Figure 40C shows exactly the opposite situation, in which a cell having a dilute intracellular fluid is placed in a concentrated extracellular fluid (*hypertonic solution*). Water passes by osmosis out of the cell, decreasing the intracellular fluid volume and increasing the extracellular fluid volume. This also concentrates the intracellular fluids while diluting the extracellular fluids. Therefore, within a few seconds, the concentrations of the two fluids reach equilibrium, but in the meantime the cell has decreased in size as shown in Figure 40D. To illustrate how rapidly osmosis can take place at the cell membrane, a red blood cell placed in pure water gains an amount of water

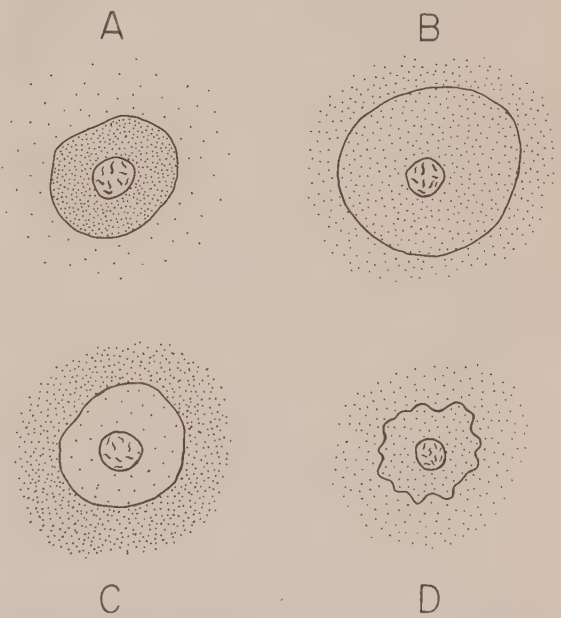


Figure 40. (A) A cell suddenly placed in a hypotonic solution. (B) The same cell after osmotic equilibrium has been established, showing swelling of the cell and equilibration of the concentration of extra- and intracellular fluids. (C) A cell placed in a hypertonic solution. (D) The same cell after osmotic equilibrium has been established, showing shrinkage of the cell and equilibration of fluid concentrations.

equal approximately to its own volume in 2 to 3 seconds. Therefore, one would expect almost complete osmotic equilibrium to take place within 15 to 20 seconds.

Isotonicity, hypertonicity, and hypotonicity of solutions. A solution that, when placed on the outside of cells, causes no osmosis through the cell membrane in either direction is said to be *isotonic* with the body fluids. For instance, a 0.9 per cent solution of sodium chloride is isotonic. On the other hand, a solution that causes osmosis of fluid out of the cell into the solution is said to be *hypertonic*. Thus, a sodium chloride solution having a concentration greater than 0.9 per cent is hypertonic. Finally, a solution that causes osmosis into cells is *hypotonic*, for instance, a sodium chloride solution of less than 0.9 per cent concentration. In other words, cells placed in an isotonic solution will maintain their volumes at a constant level, cells placed in a hypertonic solution will shrink, and cells placed in a hypotonic solution will swell.

Effect on the intra- and extracellular fluids of injecting water into the circulation. When water is injected into the blood, it is rapidly mixed by blood flow and diffusion, mixing everywhere in the extracellular fluids within 10 to 20 minutes. And, within a few seconds after the water has come into contact with the cells, it passes by osmosis through the cell membrane to mix with the intracellular fluids as well. Therefore, at the end of 10 to 20 minutes, the water has become distributed almost evenly throughout the body, both extracellularly and intracellularly.

Effect of dehydration. Dehydration means loss of water from the body, by evaporation from the body surface or from the lungs or by excretion from the body in the form of very dilute urine or in other ways. At any rate, the effects are precisely the reverse of those that occur when water is injected. The water is actually lost from the extracellular fluid compartment, but because the extracellular fluid remains in osmotic equilibrium with the intracellular fluid, water passes from the cells by osmosis into the extracellular fluid almost immediately, bringing the concentrations of the two fluids back to equilibrium. Therefore, in essence, loss of water from the body by any means causes proportionate loss from both extracellular and intracellular fluids.

Effect of administering isotonic salt (sodium chloride) solution. When sodium chloride solution is injected intravenously, or when a person absorbs such a solution from his gastrointestinal tract, it passes directly into the extracellular fluids. Because the sodium and chloride ions do not go through the cell pores, they oppose osmosis of water into the cells. Therefore, if the osmotic concentration of the administered fluid is exactly equal to that of the normal extracellular fluids, none of the administered fluid will go into the cells, but will simply increase the volume of the extracellular fluid.

MEASUREMENT OF FLUID VOLUMES OF THE BODY

The dilution principle. Frequently it is important for the physiologist to measure the fluid volumes in different compartments of the body. To do this, he utilizes a basic principle called the *dilution principle* as follows:

Figure 41 illustrates a fluid chamber in which a small quantity of dye or other foreign substance is injected. After the substance mixes with the fluid in the chamber, its concentration becomes very dilute and equal in all areas of the chamber (Fig. 41B). Then a sample of the fluid is removed and the concentration of the substance analyzed by chemical, photoelectric, or any other means. The volume of the chamber can then be calculated by the following formula:

$$\text{Volume in ml.} = \frac{\text{Quantity of test substance injected}}{\text{Concentration per ml. of sample fluid}}$$

It should be noted that all one needs to know is (1) *the total quantity of the test substance* injected into the chamber and (2) the *concentration of the substance in the fluid after dispersement.*

Measurement of total body water. To measure total body water one can inject either *heavy water* or *radioactive water* into a person and then 30 minutes to an hour later remove a sample of blood and measure the concentration of heavy water or radioactive water in the blood or plasma. Since either one of these types of water disperses throughout the body in exactly the same way as normal water, one can use this measurement to calculate the total body water. Let us assume that 50 ml. of heavy water has been injected into the person and that its concentration in each ml. of plasma is found to be 0.001 ml. Dividing this concentration into the total amount injected (using the above formula), we find that the total body water of this person is 50 liters, which is a value slightly higher than that of the normal adult man.

Measurement of extracellular fluid volume. To measure the extracellular fluid volume we

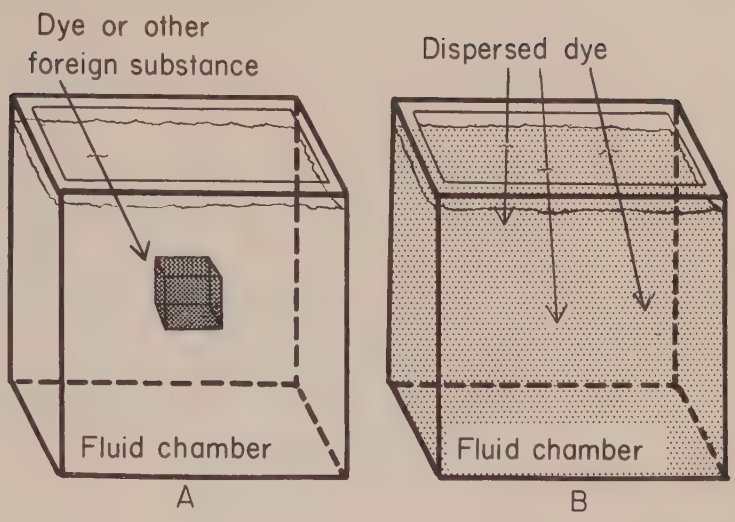


Figure 41. The dilution principle for measuring the volume of a fluid compartment.

simply use a different test substance, a substance that will diffuse everywhere in the extracellular fluid compartment but will not go through the cell membranes into the intracellular fluid compartment. Substances of this type include *radioactive sodium, thiocyanate ions,* and *inulin,* all of which have been used at one time or another for determining the extracellular fluid volume. The procedure is almost identical with that used for measuring total body water. The test substance is injected intravenously, and a sample of plasma is removed about 30 minutes later, after appropriate mixing throughout the entire extracellular compartment has occurred. The same formula above is used to calculate the extracellular fluid volume.

CALCULATION OF INTRACELLULAR FLUID VOLUME. Once the total body water has been determined using heavy water or radioactive water and the extracellular fluid volume has been determined using radioactive sodium or one of the other test substances, one can calculate intracellular fluid volume by subtracting extracellular fluid volume from total body water.

Measurement of plasma volume. The plasma volume is measured by injecting some substance that will stay in the plasma compartment of the circulating blood. Such a substance frequently used is a dye called T-1824 or Evans blue which, upon intravenous injection, combines almost immediately with the plasma proteins. Since plasma proteins do not leak readily out of the plasma compartment of the blood, neither will the dye. After appropriate mixing of the blood has occurred, within about 10 minutes, a sample of blood is removed, the plasma separated from the red cells, and the concentration of dye in the plasma measured. Then using the preceding formula, one determines the plasma volume.

CALCULATION OF INTERSTITIAL FLUID VOLUME. Once the plasma and extracellular fluid volumes have been measured, one can calculate the interstitial fluid volume by subtracting plasma volume from extracellular fluid volume.

CALCULATION OF BLOOD VOLUME. If one knows the relative percentage of the blood that is blood cells, he can calculate the total blood volume from the plasma volume. To determine the percentage of the blood that is plasma, blood is centrifuged at a high speed for 15 to 30 minutes, which separates the plasma from the cells so that the relative percentages of these two can be measured directly. The percentage of blood cells is called the *hematocrit* of the blood. From the hematocrit and the measured plasma volume one calculates total blood volume using the following formula:

$$\text{Blood volume} = \frac{100}{100 - \text{Hematocrit}} \times \text{Plasma volume}$$

REFERENCES

Adolph, E. F., and others: *Physiology of Man in the Desert.* New York, Interscience Publishers, Inc., 1947.

Drinker, C. K.: *The Lymphatic System.* Stanford University Press, 1942.

Elkinton, J. R., and Danowski, T. S.: *The Body Fluids.* Baltimore, The Williams & Wilkins Company, 1955.

Giebisch, G.: Kidney, water and electrolyte metabolism. *Ann. Rev. Physiol.* 24:357, 1962.

Robinson, J. R.: Metabolism of intracellular water. *Physiol. Rev.* 40:112, 1960.

Ullrich, K. J., and Marsh, D. J.: Kidney, water, and electrolyte metabolism. *Ann. Rev. Physiol.* 25:91, 1963.

Wirz, H.: Kidney, water, and electrolyte metabolism. *Ann. Rev. Physiol.* 23:577, 1961.

CAPILLARY MEMBRANE DYNAMICS, AND THE BODY'S SPECIAL FLUID SYSTEMS

6

CAPILLARY MEMBRANE DYNAMICS

Molecules diffuse continually both inward and outward through the capillary pores. Ordinarily, the rates of diffusion in the two directions are almost exactly equal so that the volumes of plasma and interstitial fluid remain essentially constant. Yet, under special circumstances the volume of fluid diffusing in one direction becomes greater than in the other direction, in which case the plasma and interstitial volumes change accordingly. Therefore, it is important that we consider the dynamics of this net exchange of fluid between the plasma and the interstitial fluid and the mechanisms by which normal volumes are ordinarily maintained.

Anatomy of the capillary system. Figure 42 illustrates a typical capillary bed, showing blood entering the capillaries from the arteriole, passing through the metarteriole, then into the capillaries, and finally to the venule. The *arteriole* has a muscular coat that allows it to contract or relax in response to stimuli which come mainly from the sympathetic nerves. The *metarteriole* has sparse muscle fibers, and the openings to the *true capillaries* are usually guarded by small *muscular precapillary sphincters* which can open and close the

entryway into the capillaries. The muscles of the metarterioles and the precapillary sphincters are controlled mainly in response to local conditions in the tissues. For instance, lack of oxygen allows these muscles to relax, which increases the flow of blood through the capillary bed and in turn increases the amount of oxygen available to the tissues.

Pressures in the capillary system. Though the pressure at the arterial end of a capillary varies tremendously depending upon the state of contraction or relaxation of the arteriole, metarteriole, and precapillary sphincters, the normal mean pressure at the arterial end of the capillary is probably between 20 and 35 mm. Hg. In the present discussion we will use an average value of about 26 mm. Hg. Likewise, the average pressure at the venous end of the capillary is about 10 mm. Hg, and the average mean pressure in the capillary is probably about 18 mm. Hg. Unfortunately, we do not know these values exactly, for they have never been determined entirely satisfactorily. Pressures have been measured at different points in capillaries with minute pipets, these pressures having been slightly higher than those given above. However, the use of the pipet itself obstructs the capillary and can cause abnormal pressures. The values given above have been estimated from experiments in which the pressures in the arterial and venous ends of a tissue bed have been changed to many different values and the rates of fluid transudation through the capillaries measured by special techniques. From these data one can calculate the values for capillary pressure.

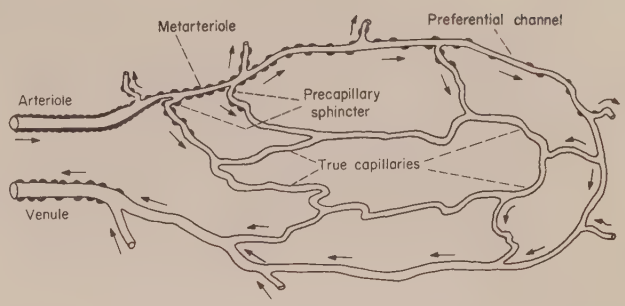


Figure 42. A typical capillary system. (Drawn from a figure by Zweifach: *Factors Regulating Blood Pressure,* Josiah Macy, Jr., Foundation, 1950.)

Tissue Pressure

Tissue pressure is the pressure of the fluid in the spaces between the cells and, therefore, is also the pressure of the fluid outside the capillaries pressing against the capillary membrane.

Tissue pressure is another value that has been difficult to measure because the interstitial spaces generally have a width of less than 1 micron, and introduction of any needle or pipet to measure the pressure can cause an abnormal reading. However, during the past few years we have studied this problem in our own laboratory in a completely different way, by implanting into tissues perforated capsules such as the one illustrated in Figure 43. The tissue grows inward through the holes and lines the inner wall of the capsule. A blood vascular system also develops in the new tissue. Yet, a large fluid space remains in the middle of the capsule. A needle connected to a manometer system can then be inserted through one of the perforations into the cavity and the pressure measured. Tissue pressures measured in this manner have been about −6 mm. Hg—that is, 6 mm. Hg less than atmospheric pressure—which demonstrates that the tissue spaces actually have a partial vacuum in them.

Measurements of tissue pressure using minute needles inserted in the tissues under special conditions have also given negative values of about −6 mm. Hg.

Pressure gradient across the capillary membrane. If the average mean pressure in the capillary is 18 mm. Hg and the average pressure outside the capillary is −6 mm. Hg, the total pressure gradient between the two sides of the membrane would be 18 − (−6) or a total of 24 mm. Hg. That is, the pressure inside the capillary

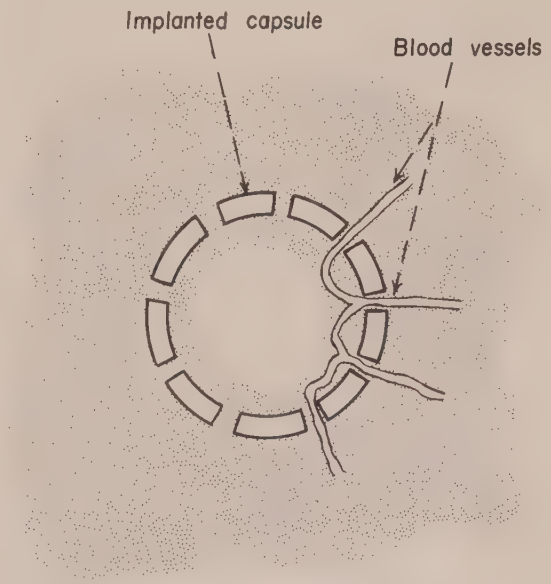


Figure 43. Diagram of the implanted capsule used to measure tissue pressure.

is 24 mm. Hg greater than the pressure outside the capillary. This pressure difference between the two sides of the membrane makes fluid tend to diffuse out of the capillary into the tissue spaces more rapidly than in the opposite direction. Fortunately, however, *colloid osmotic pressure,* another mechanism operating at the capillary membrane, opposes this tendency for fluid to leak out of the capillaries, as is discussed in the following section.

Colloid Osmotic Pressure at the Capillary Membrane

The principle of osmotic pressure. The principle of *osmosis* was discussed in the previous chapter. Now let us discuss an effect of osmosis called *osmotic pressure.* Figure 44 illustrates a U tube with a membrane at the bottom. In chamber A is a solution of water and a non-diffusible solute, while in chamber B is pure water. One can readily see that osmosis will take place at the membrane, causing a large quantity of water to diffuse from chamber B into chamber A. However, after a period of time the water level will have fallen in chamber B and the solution level in chamber A will have risen, creating a pressure difference between the two chambers. Eventually, the pressure difference becomes so great that it completely stops all osmosis, because the pressure itself pushes water molecules in the opposite direction through the membrane to oppose the osmosis. The pressure at this point is the osmotic pressure.

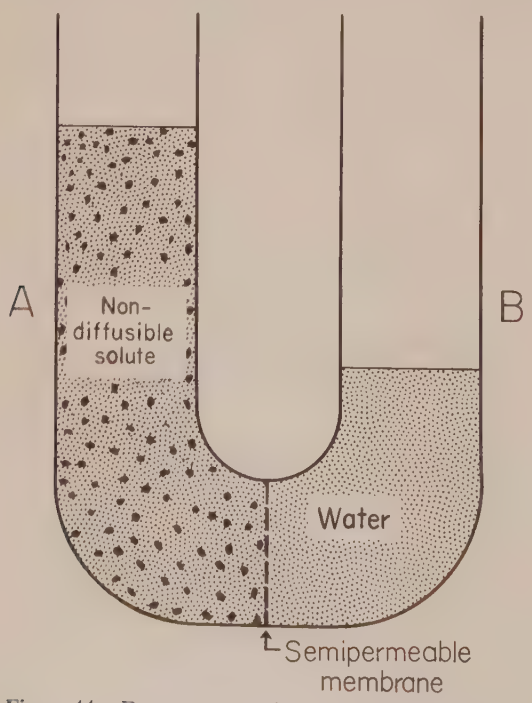


Figure 44. Demonstration of osmotic pressure on the two sides of the semipermeable membrane.

If a red blood cell were suddenly dropped into pure water, the presence of large quantities of non-diffusible solutes inside the cell would be great enough to cause an osmotic pressure at the red cell membrane equal to over 5000 mm. Hg. This 5000 mm. Hg causes very rapid influx of water into the cell, as was pointed out in the previous chapter, causing the cell to swell to double its normal volume in 2 to 3 seconds. Fortunately, however, the red blood cell is normally bathed in plasma which contains almost identically the same quantity of non-diffusible solutes as the cell. Therefore, at the normal red cell membrane there is no osmotic pressure at all.

Colloid osmotic pressure. The amount of osmotic pressure that develops at a membrane depends on the concentrations of *non-diffusible substances* on the two sides of the membrane. Furthermore, the relative sizes of the dissolved substances and of the pores determine which substances will be diffusible and which will be non-diffusible. For instance, at the cell membrane almost every constituent of the body fluids besides water is non-diffusible. However, at the capillary membrane, where the pore diameters are 10 times as large as at the cell membrane (70 angstroms in contrast to 7 angstroms), almost all constituents in the extracellular fluids can pass through the pores, with the single exception of the plasma proteins. Therefore, at the capillary membranes the only substance that causes osmosis and osmotic pressure is the *proteins* in the plasma and interstitial fluid.

Plasma normally has a concentration of proteins of about 7 grams per cent, while interstitial fluid normally has a concentration of about 2 grams per cent, making a difference between the two sides of about 5 grams per cent. As a result, osmosis of fluid should tend to occur always from the interstitial fluid into the capillaries. The pressure gradient that develops across the capillary membrane as a result of these osmotic forces is called *colloid osmotic pressure,* or by some physiologists *oncotic pressure.* The reason for the term *colloid* osmotic pressure is that protein solutions were formerly believed to be colloidal solutions rather than true solutions as we now know them to be.

COLLOID OSMOTIC PRESSURE OF PLASMA AND INTERSTITIAL FLUIDS. If pure plasma were placed on one side of a capillary membrane and pure water on the other side, the colloid osmotic

pressure that would develop would equal on the average 28 mm. Hg. If pure interstitial fluid were placed on one side of the membrane and pure water on the opposite side, the average colloid osmotic pressure that would develop would be 4 mm. Hg. Thus we say that the colloid osmotic pressure of plasma is 28 mm. Hg and of interstitial fluid 4 mm. Hg.

The *colloid osmotic pressure gradient* between the two sides of the membrane is equal to the difference between the colloid osmotic pressures of the two fluids on the respective sides of the membrane. Therefore, the colloid osmotic pressure gradient at the capillary membrane is 28 − 4, or 24 mm. Hg.

Equilibration of Pressures at the Capillary Membrane—The Law of the Capillaries

One will note from the preceding discussion that under normal conditions the fluid pressure gradient across the capillary membrane (24 mm. Hg) is exactly equal to the colloid osmotic pressure gradient across the membrane (24 mm. Hg). However, the fluid pressure gradient tends to move fluid out of the capillary, while the colloid osmotic pressure gradient tends to move fluid into the capillary. This balance between the two forces obviously explains how it is possible for the circulation to keep its blood volume constant even though the capillary pressure is considerably higher than the tissue pressure. Were it not for the colloid osmotic forces, fluid would be lost continually from the circulation until eventually the blood volume would be insufficient to maintain cardiac output.

The normal state of equilibrium between the forces tending to make fluid leave the capillaries and those tending to return fluid to the capillaries is called the *law of the capillaries,* and it is illustrated mathematically in Figure 45. The mean fluid pressure in the capillary is 18 mm. Hg and the tissue pressure is −6 mm. Hg making a total fluid pressure of 24 mm. Hg tending to move fluid out of the capillary. On the other hand, the colloid osmotic pressure in the capillary is 28 mm. Hg and in the interstitial fluid 4 mm. Hg, making a net gradient again of 24 mm. Hg. Thus, the two pressures are mathematically in balance so that the fluid volumes of neither the plasma nor interstitial spaces will be changing.

Effect of non-equilibrium at the capillary membrane. Occasionally the forces at the capillary membrane lose their state of equilibrium because one or more of them changes to a new value. When this occurs, fluid transudes through

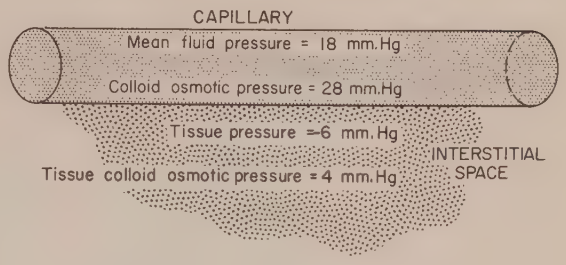


Figure 45. Mean capillary dynamics.

the membrane very rapidly and causes a new state of equilibrium to develop, usually within a few minutes to an hour or so. The basic principles of this reestablishment of equilibrium are essentially those described in the previous chapter for reestablishment of equilibrium between fluids outside and inside cells.

Let us consider, for example, an increase in capillary pressure from the normal value of 18 up to 25 mm. Hg. This would cause a net force across the capillary membrane of 7 mm. Hg in the outward direction, which would cause rapid transudation of fluid into the tissue spaces. The loss of fluid out of the plasma would cause the *capillary pressure to fall* because of decreasing blood volume and would cause the *plasma colloid osmotic pressure to rise* because of increasing concentration of the plasma proteins. In addition, the *tissue pressure would rise* because of the increasing volume of interstitial fluid, and the *tissue colloid osmotic pressure would fall* because of dilution of the tissue proteins. Thus, after a short period of time, we would have a capillary pressure of 22 mm. Hg, a plasma colloid osmotic pressure of 29 mm. Hg, a tissue pressure of −4 mm. Hg, and a tissue colloid osmotic pressure of 3 mm. Hg. If the student will add these pressures, he will see that a new state of equilibrium has been established.

In a similar manner, any other change in any one of the pressure values on the two sides of the membrane will cause rapid osmosis of fluid through the capillary membrane until equilibrium is reestablished within a few minutes to an hour or more.

Flow of Fluid Through the Tissues

In addition to the very rapid diffusion of water and dissolved substances through the capillary membrane, there is a small amount of actual *flow* of fluid through the tissues. The distinction between diffusion and flow is the following: Diffusion means movement of each molecule along its own pathway irrespective of all other mole-

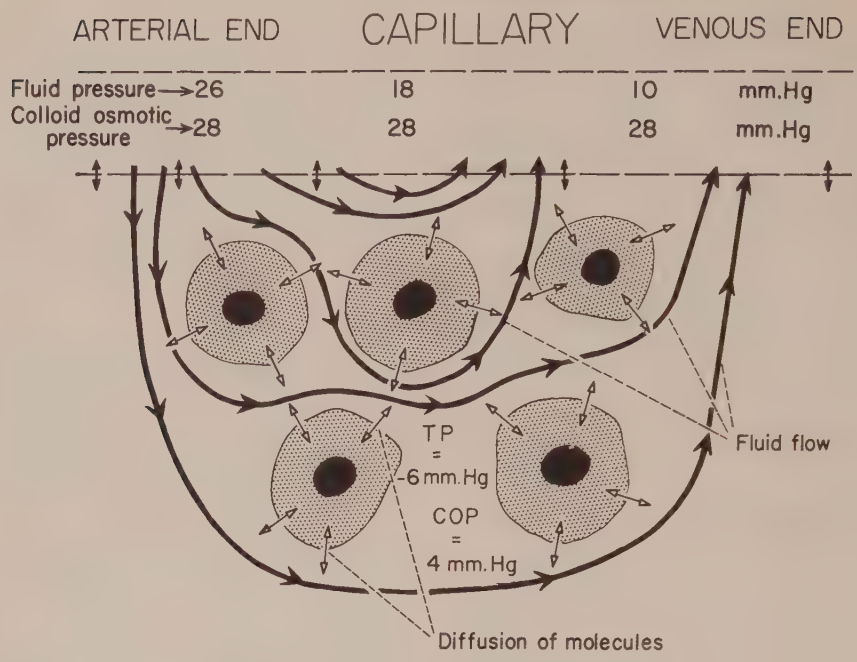


Figure 46. Flow of fluid through the tissue spaces.

cules, whereas flow means that large quantities of molecules are moving in the same direction.

Flow of fluid through the tissues is caused by unequal dynamics of fluid transfer at the arterial and venous ends of the capillaries. If we observe Figure 46, we will see that the fluid pressure in the capillary at the arterial end is 26 mm. Hg, while at the venous end it is 10 mm. Hg. Now, let us study the forces that cause fluid movement through the capillary membrane at both ends of the capillary. A sum of all the forces acting at the arterial end of the capillary gives a net pressure of 8 mm. Hg in favor of movement of fluid out of the capillary, which is called the *filtration pressure* at the arterial ends of the capillaries.

At the venous ends of the capillaries, the greatly decreased fluid pressure in the capillary causes a balance of forces in the opposite direction. This time the sum of the forces is 8 mm. Hg tending to move fluid inward rather than outward. This pressure gradient is called the *absorption pressure.*

Note that the filtration and absorption pressures are equal, so that approximately equal amounts of fluid pass out of the arterial end of the capillary and return by the venous end. A small portion of the fluid, however, between $\frac{1}{10}$ and $\frac{1}{100}$ of the total, returns to the circulation by way of the lymphatics rather than by absorption into the venous ends of the capillaries.

THE LYMPHATIC SYSTEM

In addition to the blood vessels, the body is supplied with an entirely separate set of very small and thin-walled vessels called *lymphatics.* These originate in nearly all the tissue spaces as very minute *lymphatic capillaries* which coalesce into progressively larger and larger lymphatic vessels, finally terminating in the neck as shown in Figure 47. The lymphatics from the entire lower part of the body, from the left arm, and from the left side of the head empty into the venous system at the juncture of the left internal jugular and subclavian veins, while the lymphatics from the right arm, right upper chest, and right side of the head empty into the veins at the juncture of the right internal jugular and subclavian veins.

The lymphatics are an accessory system for flow of fluid from the tissue spaces into the circulation. The lymphatic capillaries are so permeable that even very large particles and protein molecules can pass directly into them along with tissue fluids. Therefore, in effect, the fluid that flows up the lymphatics is actually overflow fluid from the tissue spaces; it is known as *lymph,* and it has identically the same constituents as normal interstitial fluid.

At many points along the lymphatics, particularly where several smaller lymphatics combine to

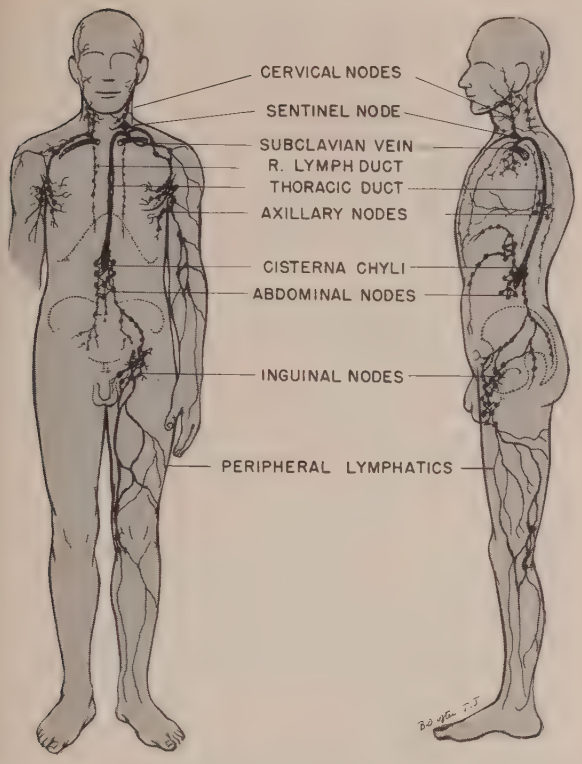


Figure 47. The lymphatic system.

form larger ones, the vessels pass through *lymph nodes,* which are small organs that filter the lymph, taking all particulate matter out of the fluid before it empties into the veins.

Functions of the Lymphatics

Return of proteins to the circulation. Undoubtedly the single most important function of the lymphatics is to return proteins to the circulation after they leak out of the capillaries. In our previous discussion of capillary dynamics we made the assumption that the plasma proteins always remain in the circulation. However, some of the pores in the capillaries are too large to prevent leakage of protein molecules, so that there is a continual small rate of protein loss, amounting each hour to approximately ½₅ of the total protein in the circulation. If these proteins were not returned to the circulation, the person would die within 12 to 24 hours. Furthermore, no other means is available by which proteins can return to the circulation except by way of the lymphatics. Later in the chapter, when we discuss edema, we will see how serious it is for proteins not to be returned to the circulation even from a single part of the body.

REGULATION OF TISSUE COLLOID OSMOTIC PRESSURE BY THE LYMPHATICS. The transport of proteins from the interstitial fluid spaces by way of the lymphatics keeps the interstitial fluid proteins low in concentration while keeping the plasma proteins high in concentration. Whenever the interstitial fluid protein concentration rises above normal, the tissue colloid osmotic pressure rises correspondingly and prevents reabsorption of fluids from the tissues into the capillaries, as was discussed previously in relation to the law of the capillaries. This increases the interstitial fluid volume and increases the tissue pressure, which in turn causes fluid to pass rapidly into the lymphatic capillaries. Thus, any increase in protein in the tissues causes fluid to wash automatically through the tissues into the lymphatics, and this washes the protein back into the circulation, thereby keeping interstitial fluid protein concentration and tissue colloid osmotic pressure at very low levels in comparison with those in the plasma.

ROLE OF THE LYMPHATICS IN REGULATION OF TISSUE PRESSURE. In a roundabout way the return of proteins by the lymphatics to the circulation also helps to keep the tissue pressure very low. Recent studies on lymphatic pressures indicate that the normal lymphatic pressure is negative like that in the interstitial spaces, measuring 2 to 4 mm. Hg less than atmospheric pressure. Any time the tissue pressure rises slightly above the lymphatic capillary pressure, fluid flows from the tissue spaces into the lymphatics and is carried back to the circulation.

One will immediately ask the question, how can the tissue pressure be more negative than the lymphatic pressure? The answer to this, in essence, is the following: All tissues of the body are continually moving, either as a result of pulsation of the arteries or as a result of actual muscular movement. These movements cause increases and decreases in the tissue pressure. When the tissue pressure rises to its peak values, fluid flows into the lymphatics and is trapped there because of valves present in the lymphatics. However, during the remainder of the time, the tissue pressure is far less than that in the lymph vessels. Thus, the *mean* tissue pressure is less than that in the lymphatics, though the *peak* pressure is actually greater than that in the lymphatics.

Flow of lymph along the lymphatics. Two principal factors determine the rate of lymph flow along the lymphatics: (1) the tissue pressure and (2) the degree of activity of the lymphatic pump.

Whenever the *tissue pressure* rises to a high value, fluid flows easily into the lymphatic capil-

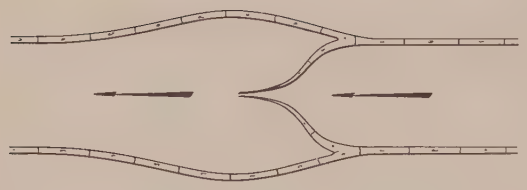


Figure 48. A lymphatic vessel, showing a lymphatic valve.

laries which have wide open communications with these spaces. Therefore, the greater the tissue pressure the greater also is the quantity of lymph formed each minute.

The *lymphatic pump,* which is the second factor affecting lymph flow, is the pumping of lymph along the lymph vessels by body movements, which may be explained as follows: Figure 48 illustrates a lymph vessel opened along one side to show a lymphatic valve. These valves are present throughout all major lymphatic channels, and every time a muscle contracts, thereby compressing the lymphatics, fluid in the vessel moves out of the compressed portion of the vessel in the direction that the valves are oriented. Since the valves are oriented toward lymph emptying sites into the veins, each compression site of the lymphatic vessel acts as a small individual pump, and the aggregate function of all these individual pumps is called the *lymphatic pump.* When one runs or exercises in any other manner, the total quantity of lymph leaving the tissues increases enormously.

In summary, the amount of lymph flowing along the lymphatics is determined (a) by the tissue pressure, which causes fluid to enter the lymphatic capillaries, and (b) by the intensity of body movements, which activates the lymphatic pump to propel lymph along the vessels.

RATE OF LYMPH FLOW. Lymph flow varies tremendously from time to time, but in the average person, the total lymph flow in all vessels is approximately 100 ml. per hour, or 1 to 2 ml. per minute. One can readily see that this small rate of lymph flow plays hardly any role in returning *fluids* from the tissue spaces to the circulation, but it does play a major role in returning *proteins* to the circulation, because average lymph contains 3 to 4 per cent protein (2 per cent from peripheral tissues and as high as 5 to 6 per cent from the liver where a major share of the lymph is formed).

Filtration of the lymph by the lymph nodes. Another function of the lymphatic system is to filter all the lymph as it returns from the tissues before emptying back into the circulation. Since large molecules or particulate matter of any type cannot be absorbed directly into the blood capillaries, it is only through the lymphatic vessels that these substances can leave the tissues. Yet, many of these can be very dangerous to the person if they get into the blood, for instance, high molecular weight protein toxins such as tetanus toxin or particles such as bacteria. Fortunately, as the lymph passes through the lymph nodes, special cells, so-called *reticuloendothelial cells,* phagocytize both foreign protein molecules and particles, then digest both of these and release the products into the lymph in the form of amino acids or other small metabolic breakdown products. Other cells in the lymph nodes form immune bodies which, as we shall see in Chapter 10, play a major role in protecting the body against foreign toxins and bacteria.

THE NORMAL "DRY" STATE OF THE BODY, AND DEVELOPMENT OF EDEMA

Importance of negative pressure in the tissues. Up to the present point very little has been said about the importance of the negative pressure in the interstitial fluid spaces. It plays two major roles: First, the negative pressure is a partial vacuum between the parts of the body and, therefore, serves to hold parts of the body together. Second, it keeps the amount of fluid in the tissues to a minimum. That is, the colloid osmotic pressure of the plasma keeps pulling fluid out of the tissue spaces until they become as small as possible, and it is in this state that the tissues normally operate. Therefore, from a relative point of view, we can say that the interstitial spaces are normally "dry," a statement that will become more meaningful as we discuss edema in the succeeding paragraphs. This is particularly beneficial

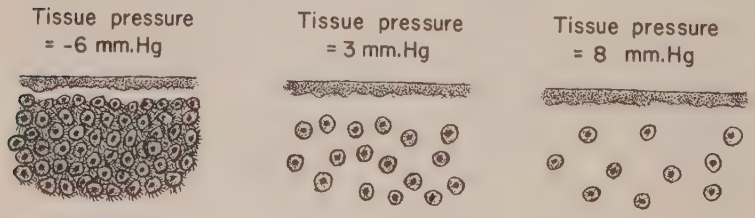


Figure 49. Effect of various tissue pressures on the amount of edema developing in the interstitial spaces.

to the diffusion of nutrients from the capillaries to the cells, for one of the basic principles of diffusion is that the rate of transport of substances by this mechanism is *inversely* proportional to the distance that the substances must diffuse.

When the body fails to maintain the "dry" state in the interstitial spaces, but instead collects large quantities of *extra* fluid, the condition is called *edema.* The transport of nutrients to the tissues becomes impaired, sometimes seriously enough to cause gangrene of the tissues. For instance, prolonged swelling of the feet frequently causes gangrenous ulcers in the skin, this resulting from too great a distance for the nutrients to diffuse from the capillaries to the tissue cells.

The Basic Causes of Edema

Positive tissue pressure. Figure 49 illustrates the compactness of the cells in normal tissues and the spreading apart of the cells when the tissue pressure rises and causes edema. Recent experiments have shown that as long as the tissue pressure remains negative, the cells remain in a compact state, but just as soon as the tissue pressure rises above the zero level and becomes positive, the cells begin to spread apart. *At tissue pressures of +3 mm. Hg, the interstitial fluid volume is often increased to as much as 3 to 4 times the normal amount,* and at +8 mm. Hg the interstitial fluid volume, at least in some tissues, often rises to as much as 20 times normal, causing very great increases in sizes of the interstitial spaces, with resultant increase in distances for nutrients to diffuse from the capillaries to the cells.

Since it is positive tissue pressure that causes edema, one can readily understand that any change in the dynamics of fluid transfer at the capillary membrane that will increase the tissue pressure from its normal negative value up to a positive value can cause edema. Three basic abnormalities of capillary membrane dynamics often result in edema. These are: (1) elevated fluid pressure in the capillaries, (2) decreased colloid osmotic pressure in the capillaries, and (3) increased tissue colloid osmotic pressure.

Elevated fluid pressure in the capillaries. Though the normal capillary pressure is about 18 mm. Hg, this can rise to values as high as 30 to 50 mm. Hg in some abnormal states of the circulation. For instance, when the *heart fails,* blood dams up in the venous system, causing back pressure on the capillaries and thereby raising the capillary pressure. Also, *blockage of the veins* can do the same thing, or *overdilatation of the arterioles* can cause such rapid blood flow into the capillaries that their pressure rises. In any of these conditions, a rise in the mean pressure in the capillaries will cause increased tissue pressure. Figure 50A shows an example in which the mean capillary pressure has risen to 33 mm. Hg. By the time enough fluid has leaked out of the circulation to cause a new state of equilibrium in accord with the law of the capillaries discussed earlier, the tissue pressure has risen to +9 mm. Hg, a pressure that causes very severe edema.

Low plasma colloid osmotic pressure. Another cause of edema is a decrease in plasma colloid osmotic pressure, which results from diminished plasma protein concentration. This occurs very frequently in persons with *kidney*

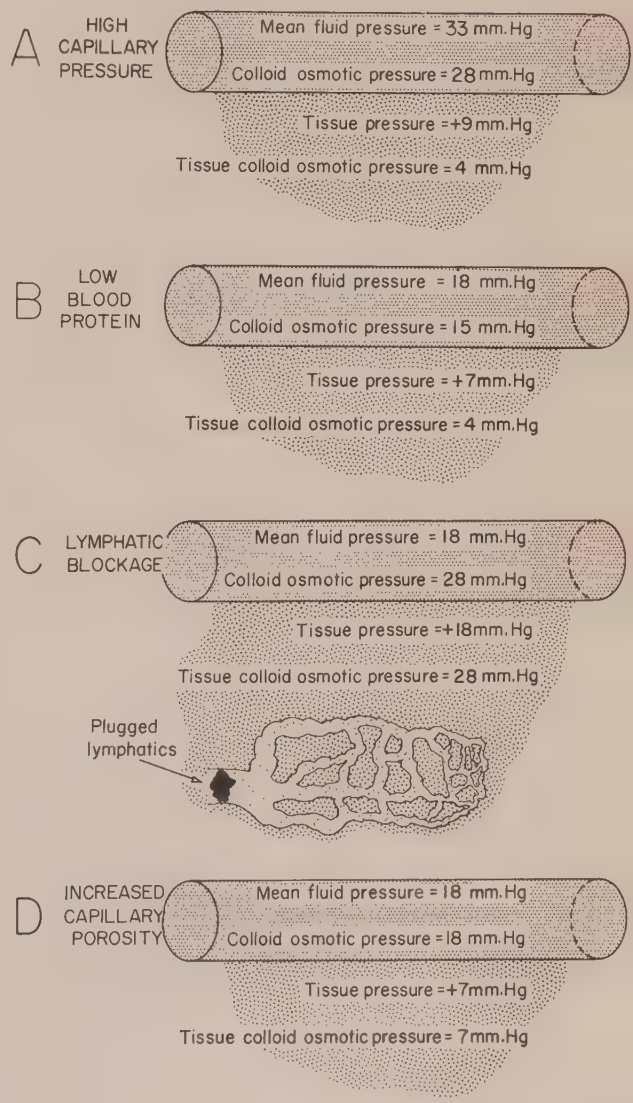


Figure 50. Various causes of interstitial fluid edema.

disease who lose tremendous quantities of plasma proteins into the urine or in persons with *severe burns* who exude large amounts of proteinaceous fluid through their denuded skin. Also, lack of adequate protein in the diet, such as occurs in *famine areas* of the world, can cause failure of formation of adequate plasma proteins. When the plasma colloid osmotic pressure falls below about 4 grams per cent, the normal negative pressure in the tissues becomes lost, and instead, the pressure rises to a positive value.

Figure 50B illustrates an instance in which the plasma colloid osmotic pressure has fallen to 15 mm. Hg. As a result, the tissue pressure has risen to +7 mm. Hg, which is high enough to cause very severe edema.

Increased tissue colloid osmotic pressure. Increased tissue colloid osmotic pressure can also cause high tissue pressure and edema. The most common cause of this is *blockage of the lymphatics,* which prevents return of proteins to the circulation. The proteins that leak through the capillary walls gradually accumulate in the tissue spaces until the tissue colloid osmotic pressure approaches the plasma colloid osmotic pressure. Such a condition is illustrated in Figure 50C, which shows a tissue colloid osmotic pressure of 28 mm. Hg, a value equal to the plasma colloid osmotic pressure. Because of this, the tissue pressure also rises to a very high value equal to the mean fluid pressure in the capillaries, 18 mm. Hg. This is a tremendously high positive pressure that can cause edema of the greatest proportions.

Lymphatic blockage commonly occurs in the South Sea Island disease called *filariasis,* in which *filariae* become entrapped in the lymph nodes and cause so much growth of fibrous tissue in the nodes that lymph flow through the nodes becomes totally or almost totally blocked. As a result, certain areas of the body, such as a leg or an arm, swell so greatly that the swelling is called "elephantiasis." A single leg with this condition can weigh as much as the entire remainder of the body, all because of the extra fluid in the tissue spaces.

INCREASED CAPILLARY POROSITY. Another condition that can increase the tissue colloid osmotic pressure is increased capillary porosity, which allows the plasma proteins to leak out of the capillaries more rapidly than the lymphatics can return them to the circulation. Therefore, proteins accumulate in the tissue fluids and increase the tissue colloid pressure, while the plasma colloid osmotic pressure is decreased because of the rapid leakage of plasma proteins, as shown in Figure 50D. The effect on the tissue pressure is an increase to +7 mm. Hg, which can cause very severe edema.

Some of the conditions that cause increased porosity are: *allergic reactions, bacterial inflammation* of local tissue areas, and damage to the capillaries resulting from *burns or other types of trauma* or *prolonged lack of blood flow* to a tissue area.

SPECIAL FLUID SYSTEMS OF THE BODY

The Potential Fluid Spaces

There are a few spaces in the body that normally contain only a few milliliters of fluid and yet can collect many liters of fluid under abnormal conditions. These spaces, called the *potential spaces* of the body, include the *intrapleural space,* which is the space between the lungs and the chest wall; the *pericardial space,* which is the space between the heart and the pericardial sac in which it resides; the *peritoneal space,* which is the space between the gut and the abdominal wall; the *joint spaces;* and the *bursae.* Ordinarily, all of these spaces are totally collapsed except for the presence of a highly viscid lubricating fluid that enables the surfaces of the spaces to slip over each other, allowing free movement of the lungs in the chest cavity, free movement of the heart in the pericardial cavity, and so forth.

Dynamics of fluid exchange. The dynamics of fluid exchange between these potential spaces and the capillaries are almost identical to the dynamics of fluid exchange at the usual capillary membrane, because the linings of these cavities are almost totally permeable to proteins and other fluids so that fluid flows with ease between the cavities and the capillaries in the surrounding tissues.

Figure 51 illustrates the intrapleural space, showing a capillary in the tissue adjacent to the

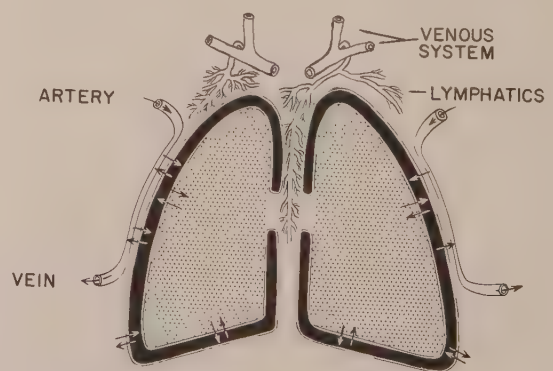


Figure 51. Dynamics of fluid exchange in the intrapleural spaces.

cavity. This figure shows diffusion of fluid back and forth between the cavity and the capillary and also diffusion of fluid through the visceral pleura into the tissue of the lungs. Obviously, with such free diffusion, the pleural space is actually nothing more than an enlarged tissue space. There is negative pressure in the intrapleural space just as there is negative pressure elsewhere in the tissues. Therefore, normally, the intrapleural space is kept "dry" in exactly the same way as the other tissue spaces of the body.

"Edema" of the potential spaces. Under abnormal conditions, a person can develop "edema" of the potential spaces in exactly the same way that he develops edema in his tissue spaces. Thus, if the capillary pressure rises too high, if the plasma colloid osmotic pressure falls too low, or if the colloid osmotic pressure in the pleural cavity rises too high, fluids will transude into the space in exactly the same way that they transude into the tissue spaces in edema. "Edema" of a potential space is called *effusion*, though physiologically it is exactly the same process as edema.

Lymphatic drainage of the potential spaces. All potential spaces have lymphatic drainage systems similar to that of the tissue spaces. Figure 51 illustrates the lymphatic drainage from the pleural cavity into the mediastinum. This drainage normally removes the proteins that leak into the intrapleural space and, therefore, keeps the colloid osmotic pressure in the space at a low value.

One of the most serious causes of effusion in a potential space is *infection*, for this usually causes the lymphatics to become plugged with large masses of white blood cells and tissue debris caused by the infection. Also, infection increases the porosity of the surrounding capillaries. Thus, the proteins cannot leave the space even though at the same time excessive quantities of proteins leak into the space. As a result, the colloid osmotic pressure increases tremendously, and, as one would expect, fluid transudes into the space in massive amounts. Thus, an infected joint swells tremendously, an infected pleural cavity develops terrific amounts of pleural effusion, and an infected abdominal cavity develops large amounts of abdominal effusion which in this case is often called *ascites*.

The Fluid System of the Eye

The fluid system of the eye maintains a constant pressure in the eyeball of almost exactly 19 mm. Hg. This pressure, along with the very strong wall of the eyeball (the sclera) keeps the eyeball in a relatively rigid shape, keeping the distances between *cornea*, *lens*, and *retina* always constant within fractions of a millimeter. This is essential to the functioning of the eye's optical system for focusing images on the retina, as will be discussed in Chapter 27.

The fluids of the eye are divided, as shown in Figure 52, into two separate compartments: the *aqueous humor*, which lies in the *anterior chamber* of the eye between the front of the lens and the cornea, and the *vitreous humor*, which lies in the *posterior chamber* between the lens and the retina. The vitreous humor is a gelatinous mass, and the fluid in it can *diffuse* through the

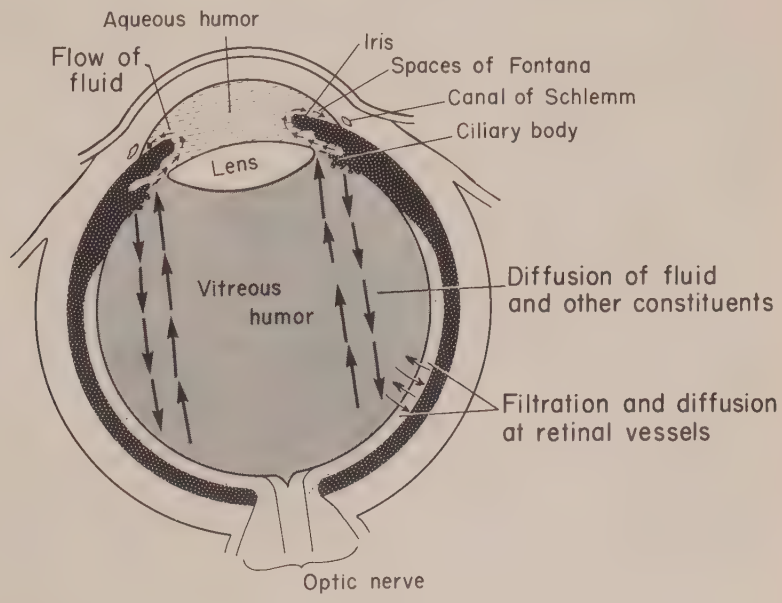


Figure 52. The fluid system of the eye.

mass but cannot *flow* through it. On the other hand, the aqueous humor in front of the lens is freely flowing. Aqueous humor also spreads backward through the ligaments of the lens along the surface of the *ciliary body* as also illustrated in Figure 52.

Formation of aqueous humor. Fluid is continually formed by the ciliary body from which it passes between ligaments of the lens and into the anterior chamber of the eye. Then it flows, as illustrated by the arrows in Figure 52, into the angle between the cornea and the iris where it makes its way through minute spaces called the *spaces of Fontana* into the *canal of Schlemm*, a circular vein that passes all the way around the eye at the juncture of the cornea and the sclera.

The ciliary body is made up of a large number of small folds of epithelium which have a total surface area in each eye of about 6 square centimeters. This epithelium secretes sodium ions continually into the aqueous humor, which in turn sets off the following effects: (1) The sodium ions cause a positive electrical charge to develop in the aqueous humor. This positive charge in turn pulls chloride and other negative ions through the epithelium, thus increasing the number of these ions in the aqueous humor as well. (2) The increased quantity of sodium ions and anions in the aqueous humor causes osmosis of water through the epithelium. Thus, indirectly, the secretion of sodium causes a continuous flow of fluid, about 2 *cubic millimeters per minute*, into the aqueous humor from the ciliary body.

Control of pressure in the eyeball. The pressure in the eye is controlled by the minute structures at the angle of the cornea and the iris. If the pressure rises too high, increased amounts of fluid leave the eye through the canal of Schlemm. On the other hand, if the pressure falls too low, the outflow of fluid from the eyeball automatically decreases. The precise mechanism by which this change in rate of fluid outflow occurs is not known, but it seems to act like a release valve on a steam boiler. When the pressure is above a critical value, the valve structures at the angle of the cornea and iris, perhaps the spaces of Fontana, appear to open up, and when the pressure is below the critical value, the spaces appear to close. Regardless of the precise details by which this mechanism works, it regulates the pressure in the eyeball almost exactly at 19 mm. Hg, rarely allowing it to rise above or fall below this value more than a few millimeters Hg.

GLAUCOMA. Glaucoma is a condition in which the pressure of the eye becomes very high, often high enough to cause blindness. The cause of the condition is failure of fluid to flow normally into the canal of Schlemm. When this happens, the continual formation of fluid by the ciliary body, without proper removal of the fluid into the canal of Schlemm, causes the eye pressure to rise sometimes to tremendous levels. For instance, *infection* or *inflammation* can cause debris or white blood cells to plug the openings of the spaces of Fontana. Also, many people develop narrowing of these spaces or pores into the canal of Schlemm for reasons that are unknown. The infectious type of glaucoma is called *secondary glaucoma* and the naturally occurring or congenital type is called *primary glaucoma.*

In rare instances the pressure in the eye rises to as high as 80 to 90 mm. Hg, though in the more usual form it rises to values of 30 to 45 mm. Hg. The very high pressures can cause blindness within a few days by damaging the optic nerve where it enters the eyeball. At this point the sclera is weak because of the many small openings through which the optic fibers must pass, and high intraocular pressure gradually pushes the nerve backward, stretching its fibers over the edge of the hole and destroying them. When the eyeball pressure is elevated only moderately, on the other hand, blindness can develop gradually over a period of years.

The Cerebrospinal Fluid System

The cerebrospinal fluid system is similar to that of the eye in that most of the fluid is formed in one area and is reabsorbed in an entirely different area. The fluid of the cerebrospinal fluid cavity plays a special role as a cushion for the brain; the density of the brain is almost the same as that of the cerebrospinal fluid, so that the brain literally *floats* in this fluid. Furthermore, the cranial vault in which the brain and cerebrospinal fluid lie is a very solid structure so that when it is hit on one side the whole vault moves as a unit, both the fluid and the brain being propelled in the same direction at the same time. Therefore, because of the cushioning effect of the cerebrospinal fluid, no damage to the brain results despite the fact that the brain is among the softest tissues of the entire body.

Formation, flow, and absorption of cerebrospinal fluid. Essentially all cerebrospinal fluid is formed by the *choroid plexuses* which are cauliflower-like growths that protrude into all four ventricles of the brain as shown in Figure 53—the two *lateral ventricles* in the two cerebral hemispheres, the *third ventricle* in the diencephalic region, and the *fourth ventricle* between the brain stem and the cerebellum. Fluid formed in

the lateral ventricles flows through two small openings called the *foramina of Monro* into the third ventricle, and fluid from the third ventricle flows through the *aqueduct of Sylvius* into the fourth ventricle. From here the fluid flows out of the ventricles, through the *foramina of Luschka* and *foramen of Magendie,* into the *subarachnoid space* which is the fluid space between the brain and the cranial vault. The fluid then flows around the brain and upward over its upper surfaces where it is absorbed into the venous sinuses through the *arachnoidal granulations.*

The choroid plexuses are similar to the ciliary processes of the eye, for they also are secretory organs, continually secreting sodium ions and water by essentially the same mechanism as that described for the eye. Likewise, the arachnoidal granulations function in a manner similar to that of the structures that lie in the angle between the iris and cornea of the eye, for these granulations act as an overflow valvular system. When the pressure in the cerebrospinal fluid is about 10 mm. Hg above that in the veins, fluid flows through the granulations, but when the pressure in the cerebrospinal fluid system is less than this level, fluid does not leave the system.

The only major difference between the fluid system of the eye and the cerebrospinal fluid system is the level of the pressure, for the pressure required to open the overflow valves in the eye is about 19 mm. Hg, while the pressure required to release fluid from the cerebrospinal fluid system into the veins is about 10 mm. Hg.

Blockage of flow in the cerebrospinal fluid system. Referring again to Figure 53, one can readily see that flow of fluid in the cerebrospinal fluid system can be blocked at many places, especially at the foramina of Monro, at the aqueduct of Sylvius, and at the arachnoidal granulations. Many babies are born with congenital blockage of the aqueduct of Sylvius so that fluid formed in the lateral and third ventricles cannot escape to the surface of the brain to be absorbed. As a result, the ventricles swell larger and larger, compressing the brain against the cranial vault and destroying much of the neuronal tissue. Another common cause of blockage is infection in the cerebrospinal fluid cavity, which causes so much debris and white blood cells in the fluid that the arachnoidal granulations become plugged. In this case, fluid accumulates in the entire cerebrospinal fluid system, and the pressure rises very high. In rare instances the pressure in the cerebrospinal fluid system can rise so high that it actually impedes blood flow in the brain so that the brain cannot receive adequate nutrition.

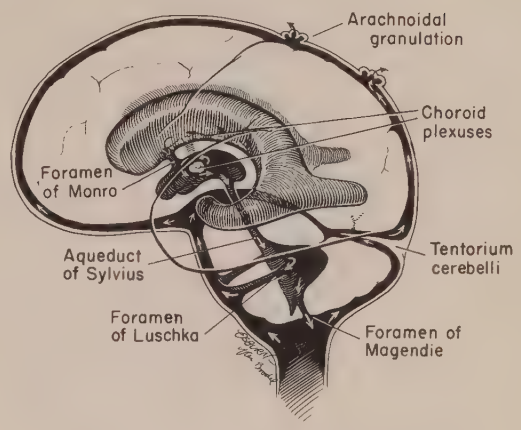


Figure 53. The cerebrospinal fluid system.

REFERENCES

Bennet, H. S., Luft, J. H., and Hampton, J. C.: Morphological classifications of vertebrate blood capillaries. *Am. J. Physiol.* 196:381, 1959.

Chambers, R., and Zweifach, B. W.: Intracellular cement and capillary permeability. *Physiol. Rev.* 27:436, 1947.

Landis, E. M.: Capillary pressure and capillary permeability. *Physiol. Rev.* 14:404, 1934.

Mayerson, H. S., Wolfram, C. G., Shirley, H. H., Jr., and Wasserman, K.: Regional differences in capillary permeability. *Am. J. Physiol.* 198:155, 1960.

Moyer, J. H., and Fuchs, M.: *Edema: Mechanisms and Management.* Philadelphia, W. B. Saunders Company, 1960.

Nicoll, P. A., and Webb, R. L.: Vascular patterns and active vasomotion as determiners of flow through minute vessels. *Angiology* 6:291, 1955.

Pappenheimer, J. R.: Passage of molecules through capillary walls. *Physiol. Rev.* 33:387, 1953.

Spector, W. G.: Substances which affect capillary permeability. *Pharmacol. Rev.* 10:475, 1958.

Ussing, H. H.: Transport through biological membranes. *Ann. Rev. Physiol.* 15:1, 1953.

Wasserman, K., Loeb, L., and Mayerson, H. S.: Capillary permeability to macro-molecules. *Circ. Res.* 3:594, 1955.

Yoffey, J. M., and Courtice, F. C.: *Lymphatics, Lymph and Lymphoid Tissue.* Cambridge, Harvard University Press, 1956.

FORMATION OF URINE BY THE KIDNEY, AND MICTURITION

7

Physiologic Anatomy of the Kidney

The kidney forms urine and in doing so regulates the concentrations of most of the substances in the extracellular fluid. It accomplishes this by removing those materials from the blood plasma that are present in excess, while conserving those substances that are present in normal or subnormal quantities.

Figures 54 and 55 illustrate respectively the gross and microscopic structures of the kidney which are responsible for this fluid purifying function. Figure 54 shows the renal artery entering the substance of the kidney and the renal vein returning from it. Urine is formed from the blood by the *nephrons*, one of which is shown in Figure 55. From these urine flows into the *renal pelvis* and then out through the *ureter* into the *urinary bladder*. The two kidneys contain approximately two million nephrons, and because each nephron operates almost exactly the same as all others, we can characterize most of the functions of the kidney as a whole by explaining the function of a single nephron.

The nephron is composed of two major parts, the *glomerulus* which filters water and solutes from the blood, and the *tubules* which reabsorb from the filtrate those substances that are needed in the body, while allowing the unneeded substances to flow into the renal pelvis as urine. The

glomerulus is a tuft of capillaries surrounded by a capsule called *Bowman's capsule.* Fluid filters out of the capillaries into this capsule and then flows from here, first, into the *proximal tubule,* second, into a long loop called the *loop of Henle,* third, into the *distal tubule,* fourth, into a *collecting tubule,* and, finally, into the *renal pelvis.* As the filtrate passes through the tubules, most of the water and electrolytes are reabsorbed into the blood, but almost all of the end-products of metabolism pass on into the urine. In this way, water and electrolytes are not depleted from the body, though the waste products of metabolism are removed constantly.

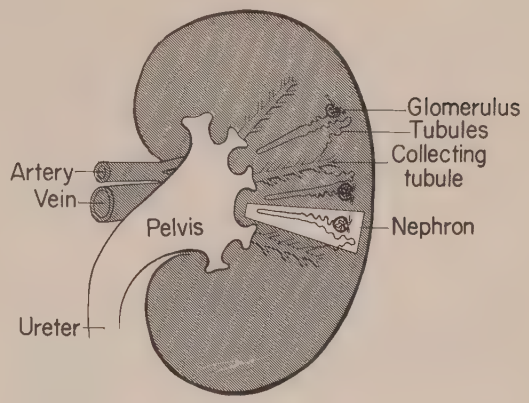


Figure 54. Principal anatomic structures of the kidney.

FUNCTION OF THE NEPHRON

In discussing the function of the nephron it is desirable to use the simplified diagram shown in Figure 56. This shows the "functional nephron" with an *afferent arteriole* supplying blood to the glomerulus and the blood leaving the glomerulus through an *efferent arteriole,* to flow into the *peritubular capillaries* and finally into the vein. Also shown are the *glomerular membrane, Bowman's capsule,* the *tubules,* and the *kidney pelvis.*

Glomerular Filtration

The membranes of the capillaries in the glomerular tuft are collectively called the *glomerular membrane.* This membrane is about as porous as the usual capillary membrane elsewhere in the body, and the same principles of fluid dynamics apply as to other capillary membranes.

However, the pressure in the glomerulus is very high, believed to be about 70 mm. Hg, in contrast to the low pressures, between 15 and 20 mm. Hg, in capillaries elsewhere in the body. Because of this high pressure, fluid leaks continually out of all portions of the glomerular membrane into Bowman's capsule. We shall see later in this chapter that most of the fluid that leaks out of the glomerular membrane is later reabsorbed from the renal tubules into the peritubular capillaries. Thus, the glomerular membrane corresponds to the arterial ends of the usual tissue capillaries where fluid leaks outward, and the peritubular capillaries correspond to the venous ends of the usual tissue capillaries where fluid is reabsorbed into the blood.

Fluid dynamics at the glomerular membrane, and the filtration pressure. The fluid pressures in the normal nephron are illustrated in Figure 57.

This figure shows that the *glomerular pressure* is normally 70 mm. Hg, while the colloid pressure in the glomerulus is normally 32 mm. Hg. The pressure in Bowman's capsule is about 20 mm. Hg, and the colloid osmotic pressure in this capsule is essentially zero. Therefore, the pressure tending to force fluid out of the glomerulus is 70 mm. Hg, and the total pressure tending to move fluid in the opposite direction is 32 + 20, or 52

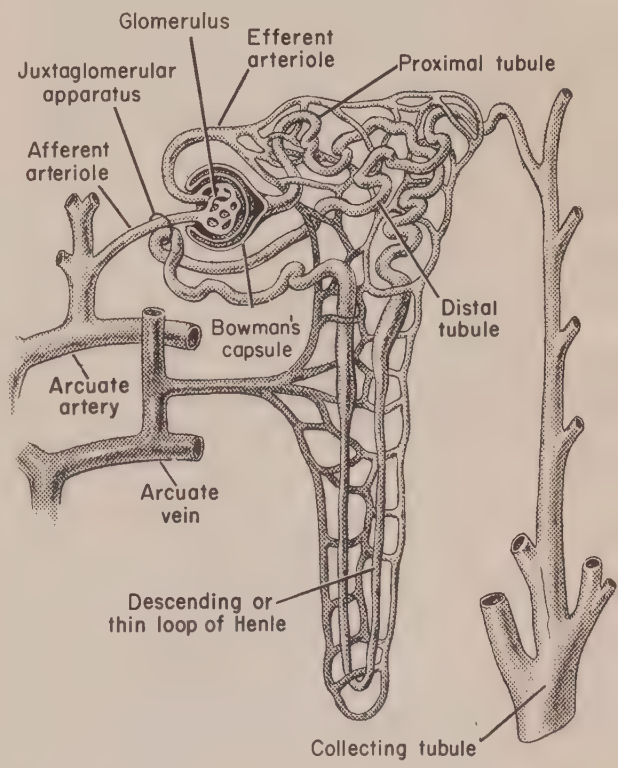


Figure 55. The nephron. (Modified from Smith: *The Kidney: Structure and Function in Health and Disease,* Oxford University Press.)

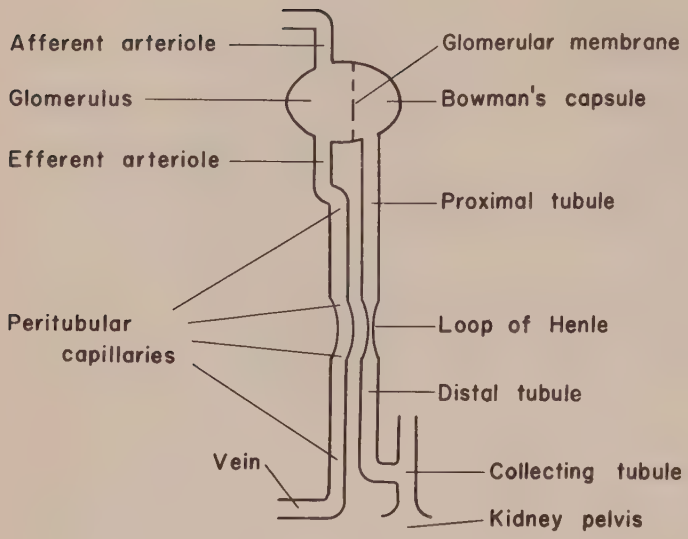


Figure 56. The functional nephron.

mm. Hg. The difference between these two, 18 mm. Hg, is the net pressure pushing fluid into Bowman's capsule; this is called the *filtration pressure.*

The rate at which fluid flows from the blood into Bowman's capsule is directly proportional to the filtration pressure. Therefore, any factor that changes any one of the pressures on the two sides of the glomerular membrane will also change the amount of glomerular filtrate formed each minute. An increase in glomerular pressure increases the rate of glomerular filtrate formation, while an increase in either the glomerular colloid osmotic pressure or the pressure in Bowman's capsule decreases the rate of filtrate formation.

EFFECT OF AFFERENT AND EFFERENT ARTERIOLAR CONSTRICTION ON FILTRATION. Figure 57B shows the effect on glomerular filtration of constricting the *afferent arteriole.* The major effect is a drastic decrease in the pressure in the glomerulus. In this figure the glomerular pressure has fallen to 49 mm. Hg, while the total of the colloid osmotic pressure plus the pressure in Bowman's capsule equals 48 mm. Hg. Therefore, the filtra-

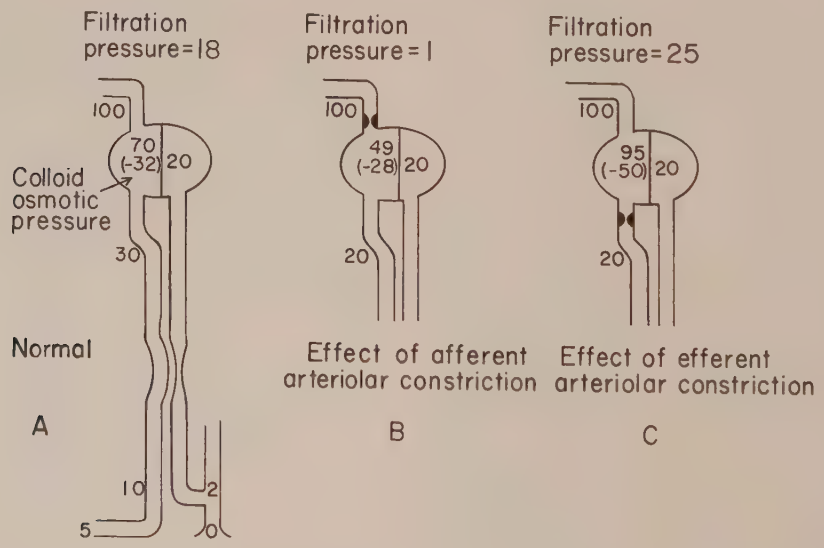


Figure 57. (A) Normal pressures at different points in the nephron, and the normal filtration pressure. (B) Effect of afferent arteriolar constriction on the pressures in the nephron and on the filtration pressure. (C) Effect of efferent arteriolar constriction on the pressures in the nephron and on the filtration pressure.

tion pressure is now only 1 mm., and, as a result, the rate of glomerular filtration has been decreased to one-eighteenth that occurring in the example of Figure 57A.

The afferent arterioles are controlled mainly by sympathetic nerves, stimulation of which constricts the arterioles and lowers the glomerular pressure, thereby decreasing glomerular filtration. On the other hand, diminished sympathetic stimulation allows afferent arteriolar dilatation and consequently greatly increased glomerular filtration.

Figure 57C illustrates the effect of *efferent* arteriolar constriction on glomerular filtration. This causes a biphasic effect. First, mild constriction greatly increases the glomerular pressure, which in turn promotes rapid filtration of fluid through the membrane. Second, intense constriction causes the blood flow through the glomerulus to decrease so greatly that almost no glomerular filtration can take place. Therefore, the quantity of filtrate formed decreases tremendously.

EFFECT OF ARTERIAL PRESSURE ON GLOMERULAR FILTRATION. When the systemic arterial pressure rises, increased amounts of blood flow into the glomerulus thereby increasing the glomerular pressure. This in turn increases the glomerular filtration rate and the output of urine. Conversely, decreasing the systemic arterial pressure decreases the urine output. This effect of arterial pressure on kidney output affords a method by which the arterial pressure itself is automatically regulated. That is, increased pressure causes increased loss of fluid from the blood. This decreases the blood and extracellular fluid volumes and thereby lowers arterial pressure. On the other hand, when the arterial pressure falls, the kidneys stop excreting fluid, thus allowing the blood and extracellular fluid volumes to increase until the arterial pressure returns to a normal value. This effect will be explained in more detail in Chapter 15 in relation to blood pressure regulation.

Characteristics of glomerular filtrate. The filtrate entering Bowman's capsule, the *glomerular filtrate,* is an ultrafiltrate of plasma. The glomerular membrane is porous enough so that water and essentially all of the dissolved constituents of plasma except proteins can filter through. Therefore, glomerular filtrate is almost identical with plasma except that only a very minute quantity of protein (0.03 per cent) is present in the filtrate, while the protein concentration in plasma is approximately 7 per cent. In Chapter 3 it was pointed out that interstitial fluid is also an ultrafiltrate of plasma, having all the constituents of plasma except proteins. Thus, glomerular filtrate is almost identical to interstitial fluid.

Approximately 180 liters of glomerular filtrate are formed each day. This is about 4.5 times as much fluid as is present in the entire body, which illustrates the magnitude of the renal mechanism for purifying the body fluids.

Tubular Reabsorption

After the glomerular filtrate enters Bowman's capsule, it passes into the tubular system where each day all but one liter of the 180 liters of glomerular filtrate is reabsorbed into the blood, the remaining one liter passing into the renal pelvis in the form of urine.

Figure 58 illustrates a cross-section of a tubular region of the kidney, showing the close proximity of the tubules to the peritubular capillaries. The tubular fluid is reabsorbed first into the interstitial spaces and then from these spaces into the capillaries. Some of the substances are reabsorbed through the tubular epithelium by the process of *active reabsorption,* while other substances are reabsorbed by the process of *diffusion* and *osmosis.*

Active reabsorption. The term *active reabsorption* means transport of substances through the tubular epithelial cells and into the interstitial spaces by means of special chemical transport mechanisms. From the interstitial spaces, the substances diffuse into the peritubular capillaries. Some of the substances reabsorbed in this manner are glucose, amino acids, proteins, uric acid, and most of the electrolytes—sodium, potassium, magnesium, calcium, chloride, and bicarbonate. Glucose, amino acids, and proteins are reabsorbed almost entirely in the proximal tubules, while electrolytes are reabsorbed in all the tubules except the collecting tubules. The active reabsorption processes for glucose, amino acids, and proteins are so powerful that ordinarily almost no glucose, proteins, or amino acids pass into the urine. On the other hand, the degree to which the

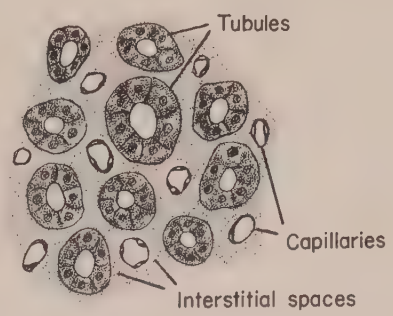


Figure 58. Cross-section of the tubules and adjacent capillaries in the kidney.

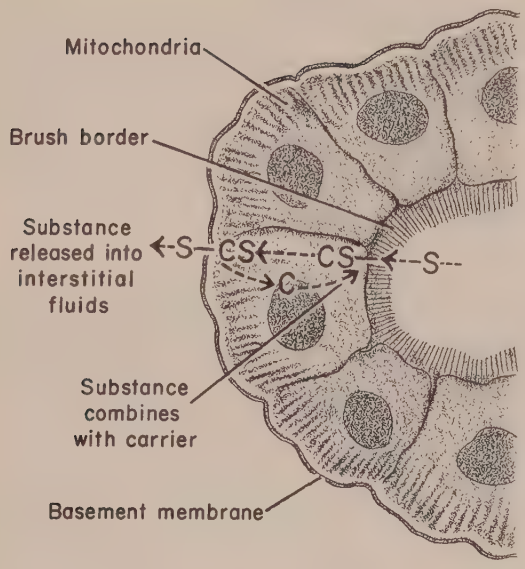


Figure 59. Mechanism of active reabsorption in the tubule.

electrolytes are reabsorbed is variable and is regulated by special control systems which are discussed in the following chapter. When an electrolyte is present in the extracellular fluids in excess, it will be reabsorbed to a less extent than when it is present in too small an amount. By this process of selective reabsorption, the renal tubules control the concentrations of electrolytes in the body fluids.

The substance that is actively reabsorbed from the tubules to the greatest extent of all is sodium chloride, the total quantity reabsorbed each day being approximately 1200 grams, which is about three fourths of all the substances actively reabsorbed from the glomerular filtrate. The reabsorption of sodium chloride is regulated by the hormone *aldosterone* which is secreted by the adrenal cortex. This regulatory mechanism is discussed in detail in the following chapter.

MECHANISM OF ACTIVE REABSORPTION. Figure 59 illustrates the mechanism of active reabsorption. Substance S is the material to be reabsorbed. The epithelial cell is impermeable to S but becomes permeable to it after it has combined with a carrier, C, at the surface of the epithelial cell. The combination CS diffuses through the cell to the opposite membrane where S splits away and passes into the interstitial fluid. Carrier C then diffuses back to the opposite pole of the cell to combine with a new quantity of S.

Several different carriers exist for transport of substances through the epithelial cells. In a few instances the same carrier transports several different substances, but in other instances there are separate carriers for separate substances. In addition, there are separate enzymes for promoting the chemical reactions between substance S and carrier C and the reactions that split S from C at the other border of the cell.

One of the most important features of active reabsorption is that it can cause absorption of a substance even when its concentration is less in the tubule than in the interstitial fluid. To do this, however, the tubular epithelial cells must expend much energy. Therefore, these cells require tremendous amounts of nutrition, and their metabolic systems are so geared that they can transform the potential energy of their nutrients into the energy required to transport substances against the concentration gradient.

One will readily recognize the similarity between the transport mechanism in the tubular epithelial membrane and that in the cell membrane as was described in Chapter 3. In both instances, specific carriers and enzymes must be available, and large amounts of energy are expended to cause the transport.

Reabsorption by diffusion and osmosis. At this point we need to recall the basic principles of diffusion and osmosis because these play a major role in the reabsorption of water and a few other substances from the tubules. Diffusion means the random movement of molecules in a fluid, and it is caused by kinetic movement of all fluid molecules. In other words, each water molecule or each dissolved molecule in the water is constantly bouncing among all the others, wending its way from place to place, going first in one direction and then another. If a suitable pore is present in a membrane, a molecule can pass through the membrane. The tubular epithelium is porous to certain types of molecules, including water molecules. Therefore, under certain circumstances water can diffuse from the tubules into the interstitial spaces of the kidney.

Osmosis is net diffusion in one direction caused by the presence of a greater concentration of nondiffusible substances on one side of a membrane than on the other side. The basic principles of osmosis were described in Chapter 5. The principal method for reabsorption of water from the tubules is by osmosis, which may be explained as follows:

Active transport of electrolytes and certain non-electrolytes such as glucose and amino acids causes the concentrations of these substances to diminish in the tubular fluids while at the same time to increase in the interstitial fluids surrounding the tubules. All of these substances are relatively non-diffusible through the pores of the membrane; therefore, a large concentration gradient develops between the two sides of the mem-

brane, which in turn causes osmosis of water through the membrane.

Thus, in effect, the initiating factor that causes reabsorption of all substances from the tubules is the active reabsorption processes, the osmotic reabsorption of water simply following in the wake of the active reabsorption of solutes.

Failure of reabsorption of unwanted substances. Some of the substances in the glomerular filtrate are undesirable in the body fluids; these substances in general are reabsorbed either not at all or very poorly by the tubules. For instance, urea, an end-product of protein metabolism, has no functional value to the body and must be removed continually if protein metabolism is to continue. This substance is not actively reabsorbed, and the pores of the tubule are so small that it diffuses through the tubular membrane several hundred times less easily than water. Therefore, while water is being osmotically reabsorbed, only a small portion of the urea diffuses out of the tubules. Instead, most of the urea remains behind and passes on into the urine. Thus, the primary function of the kidney is this separation process in the tubules, the tubules reabsorbing those substances such as amino acids, electrolytes, and water which are needed by the body while at the same time allowing urea, a substance that is not needed by the body, to pass on into the urine.

Other substances that have a fate similar to that of urea include creatinine, phosphates, sulfates, nitrates, uric acid, and phenols, all substances that are end-products of metabolism and would damage the body if they remained in the body fluids.

Active Tubular Secretion

A few substances are actively secreted from the blood into the tubules by the tubular epithelium. The substances normally secreted are creatinine, potassium, and hydrogen ions. Active secretion presumably occurs by the same mechanism as active reabsorption but in the reverse direction.

In some primitive fishes, tubular secretion is the only means by which waste products are removed. Also, a number of drugs such as penicilin, Diodrast, Hippuran, and phenolsulfonphthalein are removed from the blood primarily by active secretion rather than by glomerular filtration. However, in normal function of the kidneys, tubular secretion is important only to help in regulation of potassium and hydrogen ion concentrations in the body fluids.

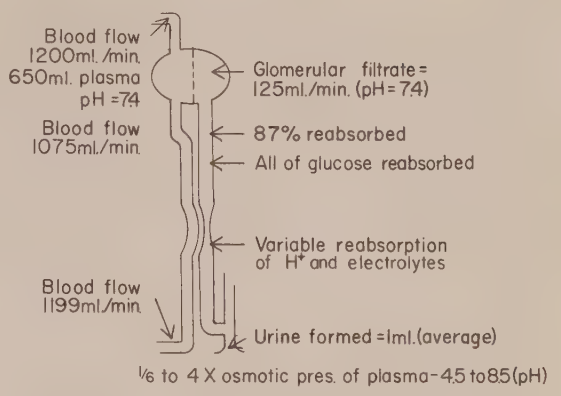


Figure 60. Overall normal function of the nephron.

Recapitulation of Nephron Function

Figure 60 shows the overall function of the nephron. The total blood flow into all the nephrons of both kidneys is approximately 1200 ml. per minute. Approximately 650 ml. of this is plasma, and about one fifth of the plasma filters through the glomerular membranes of all the nephrons into Bowman's capsules, forming an average of 125 ml. of glomerular filtrate per minute. The glomerular filtrate is actually plasma minus the proteins. The pH of glomerular filtrate is approximately 7.4, which is equal to that of the plasma. As the glomerular filtrate passes downward through the tubules, approximately 87 per cent of the water and electrolytes is reabsorbed in the proximal tubules, while all of the glucose and proteins and a major portion of the amino acids are reabsorbed. As the remaining 13 per cent of the glomerular filtrate passes through the loop of Henle, distal tubules, and collecting tubules variable amounts of the remaining water and electrolytes are absorbed, depending on the need of the body for these substances, as is discussed in the following chapter. The pH of the tubular fluid may rise or fall depending on the relative amounts of acidic and basic ions reabsorbed by the tubular walls. Also, the osmotic pressure of the tubular fluid may rise or fall depending on whether large quantities of electrolytes or large quantities of water are actively reabsorbed. Thus, the pH of the finally formed urine may vary anywhere from 4.5 to 8.2, while the crystalloidal osmotic pressure may be as little as one fourth that of plasma or as great as four times that of plasma.

The final quantity of urine formed is normally about 1 ml. per minute or $\frac{1}{125}$ of the amount of glomerular filtrate filtered each minute. This 1 ml. of urine contains about one half of the urea that is in the original glomerular filtrate, all of the

creatinine, and large proportions of the uric acid, phosphate, potassium, sulfates, nitrates, and phenols. Thus, even though almost all the water in the tubular fluid is reabsorbed, a very large proportion of the waste products in the original glomerular filtrate is never reabsorbed, but instead passes into the urine in a highly concentrated form.

THE CONCEPT OF CLEARANCE

The function of the kidney is actually to *clean* or to *clear* the extracellular fluids of the body. Every time a small portion of plasma filters through the glomerular membrane, passes down the tubules, and then is reabsorbed into the blood, it leaves behind in the tubules a large proportion of the unwanted products. In doing so it is "cleared" of these substances. For instance, out of the 125 ml. of glomerular filtrate formed each minute, approximately 60 ml. of that reabsorbed from the tubules leaves its urea behind. In other words, 60 ml. of plasma is cleared of urea each minute by the kidneys. In the same way, each minute 125 ml. of plasma is cleared of creatinine; 12 ml., of uric acid; 12 ml., of potassium; 25 ml., of sulfate; 25 ml., of phosphate; and so forth.

Calculation of renal clearance. The method by which one determines how much plasma is cleared of a particular substance each minute is to take simultaneous samples of blood and urine. From these samples the quantity of the substance in each milliliter of blood is analyzed chemically, and the quantity of the substance appearing in the urine each minute is determined. By dividing the quantity of the substance in each milliliter of plasma into the quantity passing into the urine during one minute, one can calculate the milliliters of plasma cleared per minute. That is,

$$\text{Plasma clearance} = \frac{\text{Milligrams secreted in urine per minute}}{\text{Milligrams in each milliliter of plasma}}$$

As an example, if the concentration of urea in the plasma is 0.2 mg. in each ml. and the quantity of urea entering the urine per minute is 12 mg., then the amount of plasma which would lose its urea during that minute would be 60 ml. To express this another way, the *plasma clearance* of urea is 60 ml. per minute.

Renal clearance as a test of kidney function. Since the primary function of the kidneys is to clear the plasma of undesired substances, one of the best means for testing overall kidney function is to measure the clearance of these substances. The clearance of urea has been one of the most widely used tests of renal function. Normal plasma clearance of urea is approximately 60 ml. per minute; the clearance is less than this if the kidneys are damaged, and the amount that it is depressed gives one an approximation of the degree of damage.

Another means by which kidney function can be measured is to inject a foreign substance intravenously and determine how rapidly the kidneys can clear it from the body fluids. The substance most widely used for this purpose is phenolsulfonphthalein, which is a red dye. After injection most of the phenolsulfonphthalein is cleared from the body fluids within an hour and can be recovered in the urine. If less than the normal amount is recovered in the urine, the kidneys probably are damaged, and the degree of damage can be estimated by the amount of depressed clearance.

Renal clearance as a research tool. By studying the clearance of special substances one can determine the rate of glomerular filtration per minute, the total rate of blood flow through the two kidneys per minute, the approximate number of functioning tubules, and so forth.

Figure 61 illustrates the method by which one determines the glomerular filtration rate. The clearance of the substance *inulin* by the kidneys is exactly equal to the glomerular filtration rate for the following reasons: Inulin filters through the glomerular membrane as easily as all other dissolved constituents of the plasma, so that the concentration of inulin in the glomerular filtrate is exactly equal to that in the plasma. However, inulin is not reabsorbed nor secreted even in the minutest degree by the tubules. Therefore, all of the inulin appears in the urine. In other words, all

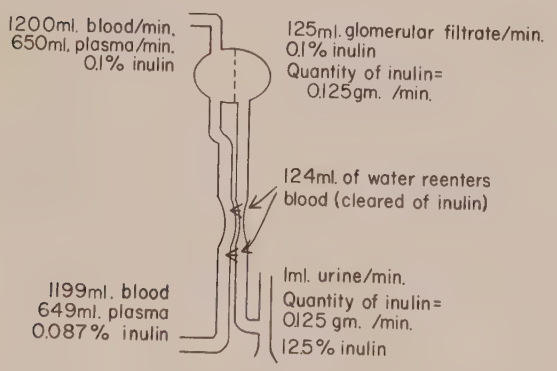


Figure 61. Measurement of glomerular filtration rate by determining inulin clearance.

of the originally formed glomerular filtrate is cleared of inulin, which means that the rate of inulin clearance is equal to the rate of glomerular filtrate formation. In Figure 61 a small quantity of inulin is injected into the blood and after mixing with the plasma its concentration is found to be 0.001 gm. in every milliliter of plasma. The amount of inulin appearing in the urine is 0.125 gm. per minute. On dividing, we find that the plasma clearance of inulin is 125 ml. per minute, which is also the amount of glomerular filtrate formed each minute.

The amount of blood flow through the two kidneys can be estimated from the clearance of either Diodrast or para-aminohippuric acid. These two substances, when injected into the blood in small quantities, are almost totally cleared by active tubular secretion. Therefore, if the plasma clearance of Diodrast is found to be 600 ml. per minute, one can be certain that at least 600 ml. of plasma flowed through the kidneys during that minute. This corresponds to a total flow of whole blood through the kidneys of 1000 ml. or more per minute.

ABNORMAL KIDNEY FUNCTION

Almost any type of kidney damage decreases the ability of the kidney to cleanse the blood. Therefore, kidney abnormalities usually cause an excess of unwanted metabolic waste products in the body fluids, and poor regulation of the electrolyte and water composition of the fluids.

Kidney shutdown. Several diseases can cause the kidneys to stop functioning suddenly and completely. Two of the most common of these are, first, poisoning of the nephrons by mercury, uranium, gold, or other heavy metals, and, second, plugging of the kidney tubules with hemoglobin following a transfusion reaction. In addition to these, almost any of the other diseases of the kidneys that are listed below can cause either gradual or rapid shutdown.

Following kidney shutdown the concentrations of urea, uric acid, and creatinine, all of which are metabolic waste products, may reach 10 times normal levels. Also, the body fluids may become extremely acidotic because of failure of the kidneys to excrete sufficient quantities of acid, and if the person continues to drink water he will become very edematous because of failure to rid himself of the ingested fluid. The person passes into coma within a few days, mainly because of acidosis. If the shutdown is complete, he will die in 8 to 14 days.

Kidney abnormalities that cause loss of nephrons. Many kidney diseases destroy large numbers of whole nephrons at a time. For instance, infection of the kidney can destroy large areas of the kidneys; trauma can destroy part of or an entire kidney; occasionally a person is born with congenitally abnormal kidneys in which many of the nephrons are already destroyed; or nephron destruction can result from poisons, toxic diseases, and arteriosclerotic blockage of renal blood vessels.

As many as three fourths of the nephrons in the two kidneys can usually be destroyed before the composition of the person's blood becomes excessively abnormal. The reason for this large margin of safety is that the undamaged nephrons can function much more rapidly than usual. The amount of blood flowing into each nephron becomes greatly increased, and the glomerular filtration rate per nephron can rise to two or more times normal. This increased activity compensates to a great extent for the lost nephrons, allowing the metabolic waste products to be removed in sufficient quantity to maintain normal body fluid composition. However, these persons usually are treading a thin line of safety because bouts of excess metabolism caused by exercise, fever, or even ingestion of too much food may present the kidneys with far more waste products than they can handle.

Obviously, as the degree of kidney destruction progresses, the derangements of the extracellular fluid become progressively more severe until finally the condition approaches that of kidney shutdown. The person then develops extreme edema, poor metabolism, and eventually coma and death.

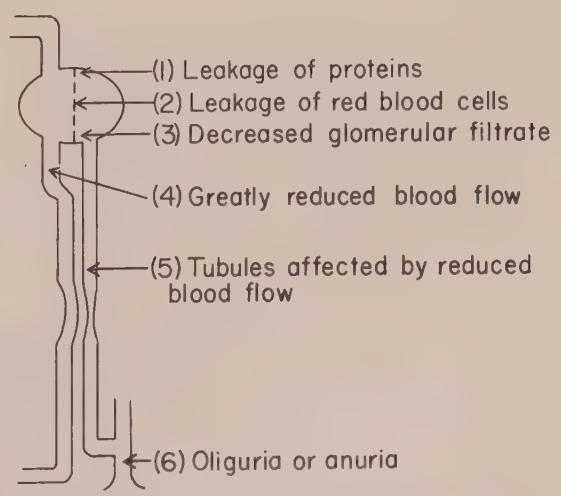


Figure 62. Effect of glomerulonephritis on the nephron.

Glomerulonephritis. A very common kidney disease is glomerulonephritis. This, like rheumatic heart disease, is an allergic disease caused by toxins of streptococcic bacteria. Almost all episodes of acute glomerulonephritis occur approximately two weeks after a severe streptococcic sore throat or some other streptococcic infection. The glomeruli become acutely inflamed, swollen, and engorged with blood. Blood flow through the glomeruli almost ceases, and the glomerular membranes become extremely porous, allowing both red blood cells and protein to flow freely into the tubules as illustrated in Figure 62. A person with glomerulonephritis has decreased kidney function, sometimes to the extent of complete kidney shutdown. If any urine is still formed, it contains large quantities of red blood cells and proteins.

In many instances of acute glomerulonephritis, the inflammation of the glomeruli regresses within two to three weeks, but even so the disease usually permanently destroys a large number of nephrons. Repeated small bouts of glomerulonephritis may destroy more and more nephrons, causing *chronic glomerulonephritis.* This can run a course of a few to many years, leading eventually to edema, depressed metabolism, coma, and death.

Effect of Diuretics to Enhance Kidney Output

A number of different drugs, called *diuretics,* can be given to a person who has either normal or diseased kidneys to increase his urine output. The actions of some of these are:

1. *Xanthines,* such as caffeine in coffee and theophylline in tea, cause dilatation of the afferent arterioles, thereby increasing the glomerular pressure and the glomerular filtration rate. As a result, the amount of urine formed per minute is greatly enhanced.

2. Administration of large quantities of *urea* or non-reabsorbed sugars such as *sucrose* can cause rapid flow of urine in the following way: These

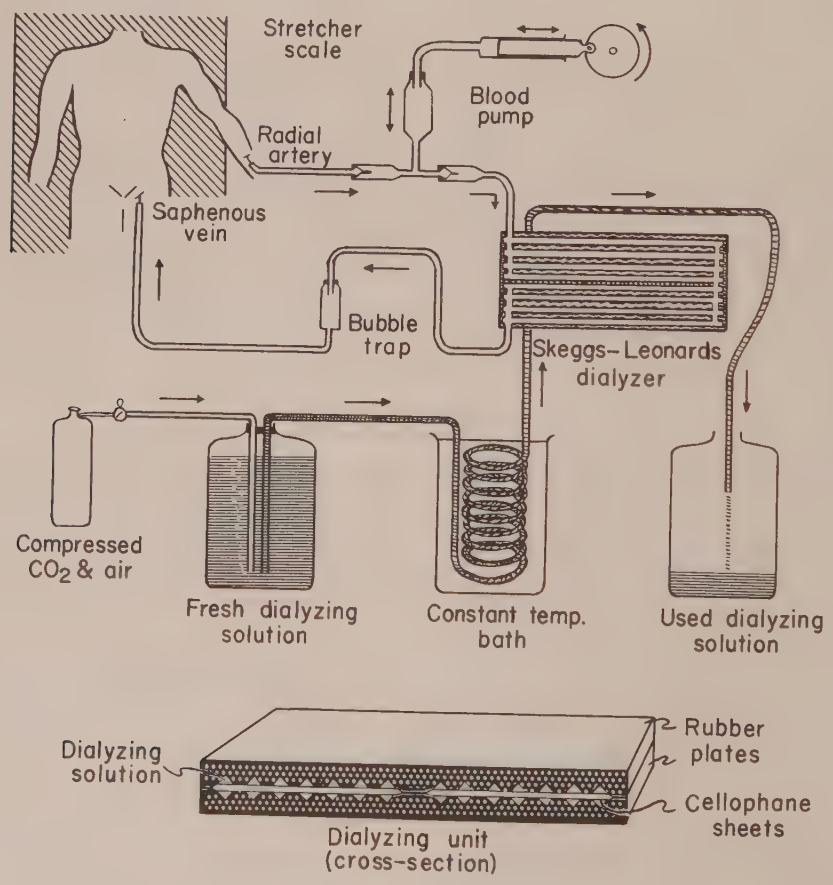


Figure 63. Diagram of an artificial kidney. (Redrawn from Bluemle, et al: *Tr. and Stud. Coll. Physicians* 20:157, 1953.)

substances are filtered through the glomerulus into the tubules, but they are not reabsorbed by the tubules to a major extent. Their presence in the tubules in excessive amounts greatly increases the crystalloidal osmotic pressure of the intratubular fluids, and thereby opposes water reabsorption. As a result, increased quantities of fluid flow on through the tubules and into the urine.

3. Certain types of drugs such as *organic mercurial compounds, chlorothiazide,* and *Diamox* either poison or block some of the specific enzymes in the tubular epithelial cells used for active reabsorption, especially those enzymes that promote reabsorption of sodium and chloride. As a result, larger quantities than usual of these substances remain in the tubules, and their crystalloidal osmotic pressure opposes the reabsorption of water, promoting increased urine flow.

Use of the Artificial Kidney

When the kidneys are damaged so severely that they can no longer maintain normal composition of the extracellular fluid, it is sometimes desirable to readjust the constituents of the extracellular fluid by use of an artificial kidney, such as the one shown in Figure 63. The artificial kidney is nothing more than a semiporous cellophane membrane arranged so that blood flows over one surface and a *dialyzing solution* flows over the other surface. The membrane is porous to all substances in the blood except the proteins and red blood cells. Almost all the dissolved substances can diffuse from the blood into the dialyzing solution, and the substances in the solution can diffuse into the blood. The fluid contains none of the waste products of metabolism. Consequently, the waste products of metabolism diffuse into the bath. On the other hand, the bath contains approximately the same concentrations of electrolytes as those found in normal plasma. Therefore, electrolytes diffuse in both directions, which prevents the blood from losing its normal electrolytes. Occasionally, the artificial kidney is used to supply supplementary nutrition. Large quantities of glucose, for instance, may be placed in the bath so that while the waste products are being removed glucose diffuses into the blood, bolstering the patient's nutritional status.

Unfortunately, the artificial kidney cannot be used with impunity because the person's blood must be rendered incoagulable during its use and because a very large amount of blood must flow through the artificial kidney to make it work. For these reasons it can be applied only for a few hours out of every several days.

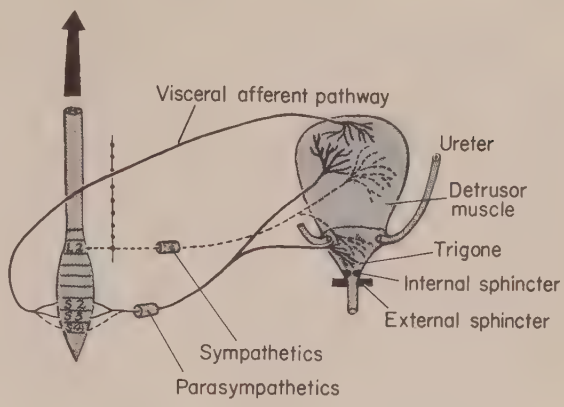


Figure 64. The urinary bladder, and nervous pathways for the micturition reflex.

MICTURITION

Transmission of urine to the bladder. Urine, formed by the kidneys, passes into the *pelves* of the kidneys, shown in Figure 54, and then through the *ureters* to the *urinary bladder,* shown in Figure 64. Urine is forced along the ureters by *peristalsis,* an intermittent wave-like constriction beginning at the pelvis and spreading downward along the ureter toward the bladder. The constriction forces the urine ahead of it. Ordinarily, the urine is transported the entire distance from the pelvis to the bladder in less than one minute.

Occasionally, severe infection or congenital abnormalities destroy the ability of the ureteral wall to contract. As a result, urine begins to collect in the kidney pelvis, causing it to swell and promoting infection that may extend into the kidney. Also, the stagnation of urine may lead to precipitation of crystalline substances, the most prominent of which are various calcium compounds that can grow eventually into large *calculi* or *renal stones* that partially or totally fill the pelvis. These stones in turn often cause extreme pain and obstruction to urine flow.

Storage of urine in the bladder. The urinary bladder is a storage reservoir designed to prevent constant dribbling of urine. The only exit of urine from the bladder is through the *urethra,* and two muscular sphincters surround the urethra, the *internal urethral sphincter* which is controlled by the autonomic nervous system and the *external urethral sphincter* which is controlled by the conscious portion of the brain. Ordinarily, both of these sphincters remain contracted so that urine cannot flow out of the bladder except when the person needs to urinate, as is explained below.

The urinary bladder itself can expand from a volume as small as one milliliter to almost a liter.

Until the bladder has filled to a volume of 200 to 400 ml., the intrabladder pressure does not increase greatly. This results from the ability of the smooth-muscle bladder wall to stretch tremendously without building up any significant tension in the muscle. However, once the bladder fills beyond 200 to 400 ml., the pressure does begin to rise, and it sometimes reaches as much as 50 mm. Hg when the bladder fills to 700 ml.

Emptying of the bladder—the micturition reflex. The term *micturition* means emptying of the bladder, and in the human being it is caused by a combination of involuntary and voluntary nervous activity which may be explained as follows: When the volume in the bladder is greater than 200 to 400 ml., special nerve endings in the bladder wall called "stretch receptors" become excited. These transmit nerve impulses through the *visceral afferent nerve pathway* into the spinal cord (see Figure 64), initiating both a conscious desire to urinate and a subconscious reflex called the *micturition reflex*. The nervous centers for the subconscious reflex are in the lower tip of the cord. The efferent impulses are transmitted from here by the parasympathetic nerves to both the bladder wall and the internal urethral sphincter. The bladder wall contracts while the internal sphincter relaxes. Then, the only impediment to urination is the still contracted external urethral sphincter. If the time and place are propitious for urination, the conscious portion of the brain will relax the external sphincter and urination will take place.

If a person wishes to urinate before a micturition reflex has occurred, he can usually initiate the reflex by contracting the abdominal wall, which pushes the abdominal contents down against the bladder and momentarily excites some of the stretch receptors in the bladder wall, thus initiating the reflex.

Many times it is not convenient to urinate when a micturition reflex takes place. In this case, the reflex usually subsides within a minute or so, and the person loses his desire to urinate. The reflex then remains inhibited for another few minutes to as much as an hour before it returns again. If the reflex is again prevented from causing urination, it will once again become dormant for another few minutes. However, as the bladder becomes more and more filled, the micturition reflex finally becomes so powerful that it is essential to urinate.

Babies, who have not developed voluntary control over the external urethral sphincter, automatically urinate every time the bladder fills. Also, adults whose spinal cords have been severed from the brain cannot keep the external sphincter contracted, so that they, too, urinate automatically when the bladder fills. However, these persons can often initiate this reflex prematurely by scratching the genital region, in this way controlling the time that the bladder will empty rather than waiting for the automatic emptying.

REFERENCES

Allen, A. C.: *The Kidney.* 2nd Ed. New York, Grune & Stratton, Inc., 1958.

de Wardener, H. E.: *The Kidney.* Boston, Little, Brown and Company, 1958.

Langston, J. B., Guyton, A. C., and Gillespie, W. J.: Acute effect of changes in renal arterial pressure and sympathetic blockade on kidney function. *Am. J. Physiol.* 197:595, 1959.

Leaf, Alexander: Kidney, water and electrolytes. *Ann. Rev. Physiol.* 22:111, 1960.

Lotspeich, W. D.: Kidney, water, and electrolyte metabolism. *Ann. Rev. Physiol.* 20:339, 1958.

Peters, L.: Renal tubular excretion of organic bases. *Pharmacol. Rev.* 12:1, 1960.

Schmidt-Nielsen, B.: Urea excretion in mammals. *Physiol. Rev.* 38:139, 1958.

Smith, H. W.: *The Kidney: Structure and Function in Health and Disease.* New York, Oxford University Press, 1951.

Smith, H. W.: *Principles of Renal Physiology.* New York, Oxford University Press, 1956.

Taggart, J. V.: Mechanisms of renal tubular transport. *Am. J. Med.* 24:774, 1958.

Winton, F. R.: *Modern Views on the Secretion of Urine.* Boston, Little, Brown and Company, 1956.

REGULATION OF BODY FLUID CONSTITUENTS AND VOLUMES

8

Now that the function of the kidneys has been explained, it is possible to discuss the mechanisms by which most of the body fluid constituents are regulated. The kidneys play a special role in the regulation of (a) electrolyte concentrations of the extracellular fluid, (b) osmotic pressure of this fluid, and (c) acidity of all the fluid. In the regulation of acidity, the respiratory system also plays a major role. All of these interrelationships are discussed.

Regulation of Sodium Ion Concentration in the Extracellular Fluid

Figure 65 illustrates the overall mechanism by which sodium ion concentration in the extracellular fluids is regulated, which is described as follows: The hormone *aldosterone* secreted by the adrenal cortex acts on the renal tubules to cause increased reabsorption of sodium. This obviously conserves the sodium in the extracellular fluids so that the extracellular fluid concentration of sodium increases. Conversely, if the quantity of aldosterone is reduced, increased amounts of sodium are lost into the urine so that the extracellular fluid concentration of sodium decreases.

The rate of secretion of aldosterone by the adrenal cortex is controlled by the sodium con-

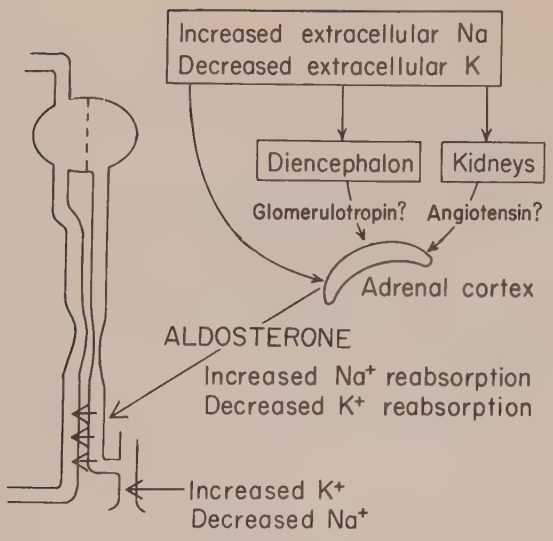


Figure 65. Postulated mechanisms for control of sodium and potassium concentration in the extracellular fluid.

centration in the extracellular fluids. The exact mechanism by which sodium exerts this effect on aldosterone secretion is unknown, but there are three different theories to explain the effect.

Some physiologists believe that sodium deficiency causes the kidneys to release a humoral factor, *angiotensin*, which then stimulates the adrenal cortex to produce aldosterone.

Others believe that sodium deficiency causes structures in the midbrain near the pineal gland to secrete another humoral substance, *glomerulotropin*, that in turn stimulates the glomerular portion of the adrenal cortex to secrete aldosterone.

Finally, it is known that sodium deficiency has a direct effect on the adrenal cortex to increase aldosterone secretion, but it is still doubtful whether this effect is enough to cause the amount of aldosterone secretion that is known to occur.

Now, let us recapitulate the overall mechanism for control of sodium concentration in the body fluids. If the sodium concentration becomes too low, the adrenal cortex secretes increased quantities of aldosterone. The aldosterone in turn promotes increased reabsorption of sodium so that the extracellular fluid concentration of sodium rises toward normal. Conversely, if the sodium concentration becomes too great, aldosterone secretion decreases, and large amounts of sodium are lost into the urine until the extracellular sodium concentration is normal again.

It is immediately recognized that this is a typical feedback regulatory mechanism. That is, the initiating stimulus, the sodium concentration, sets off a series of reactions that returns the initiating factor to its normal mean level.

Regulation of Other Electrolytes in the Extracellular Fluid

The renal tubules also have the ability to regulate the concentrations of all the other electrolytes in the extracellular fluid as well as sodium, though the mechanisms of these other regulations are not nearly so well understood.

Potassium concentration is mainly regulated as a secondary effect of the regulation of sodium concentration. That is, when aldosterone increases the reabsorption of sodium, it simultaneously decreases potassium reabsorption. Furthermore, increased potassium concentration stimulates the adrenal cortex to secrete aldosterone in the same way that low sodium concentration stimulates it. Thus, at the same time that the aldosterone mechanism helps the body to retain sodium, it simultaneously helps the extracellular fluid rid itself of potassium, thereby keeping the potassium concentration in the extracellular fluid low.

In addition to the aldosterone mechanism for potassium regulation, still another kidney mechanism helps in potassium regulation as follows: The distal tubules of the kidneys secrete potassium into the urine whenever the potassium concentration becomes too high in the extracellular fluid; this effect presumably results from a direct stimulatory effect of high potassium concentration on the tubular epithelial cells.

Chloride and *bicarbonate* ion concentrations are also regulated mainly secondarily to the regulation of sodium concentration. When sodium is reabsorbed from the tubules, this transfers positively charged ions out of the tubular fluid into the interstitial fluids of the kidneys, creating a state of electronegativity in the tubules and electropositivity in the interstitial fluids. The negativity in the tubules repels negative ions from the tubules while the positivity of the interstitial fluids attracts the negative ions. Therefore, for every sodium ion reabsorbed from the tubules, a chloride or bicarbonate ion is also reabsorbed. Sometimes more chloride than bicarbonate ions are reabsorbed while at other times more bicarbonate ions are reabsorbed; which one it will be is determined by the acid-base balance of the extracellular fluid, as is discussed later in the chapter.

The details of the mechanisms by which calcium, magnesium, and phosphate concentrations are regulated by the kidneys are not very clear. Yet, too high a concentration of any one of these

substances in the extracellular fluid causes the tubules to reject it and to pass it on into the urine. On the other hand, a low concentration causes the opposite effect, that is, rapid reabsorption of the substance until its concentration in the extracellular fluids returns to normal.

Regulation of Crystalloidal Osmotic Pressure of the Extracellular Fluid

The "counter-current" mechanism of the kidney. Before it is possible for us to explain the mechanism by which crystalloidal osmotic pressure of the extracellular fluid is regulated, we must describe the means by which the kidneys can form either a concentrated or a dilute urine. Figure 66 illustrates the so-called *counter-current mechanism* of the kidney that is responsible for the ability of the kidney to concentrate the urine. This can be explained as follows: The kidneys are divided into two major portions, the *cortex* and the *medulla.* The glomeruli and the proximal and distal tubules are located in the cortex. On the other hand, the *loop of Henle,* which is a portion of the tubule, extends deeply into the medulla. Also, the *collecting tubule* passes through the medulla before it empties into the renal pelvis. The arrangement of the *peritubular capillaries* in the medulla is also peculiar: These form long loops called the *vasa recta,* the blood passing downward into the medulla and then back upward to the cortex before emptying into the veins.

Keeping in mind the preceding anatomic details and referring again to Figure 66, we can now explain the counter-current mechanism. As the tubular fluid passes through the loop of Henle, sodium chloride is actively absorbed into the interstitial fluids of the medulla. The concentration of sodium chloride in the interstitial fluids becomes about three times as great as that of the usual extracellular fluid. The loop arrangement of the vasa recta prevents the blood from carrying large quantities of this sodium chloride away from the interstitial fluids in the following manner: As the blood flows downward in the loop, sodium chloride diffuses rapidly into the blood, its concentration rising very rapidly to a value of 750 mEq. per liter by the time the blood has reached the bottom of the loop (see Figure 66). Then, as the blood flows upward and out of the medulla, most of this sodium chloride diffuses back out of the blood into the extracellular fluid. Thus, the "counter-current" flow of blood in the vasa recta prevents excessive loss of sodium chloride out of the medulla.

Now, keep in mind the fact that the interstitial

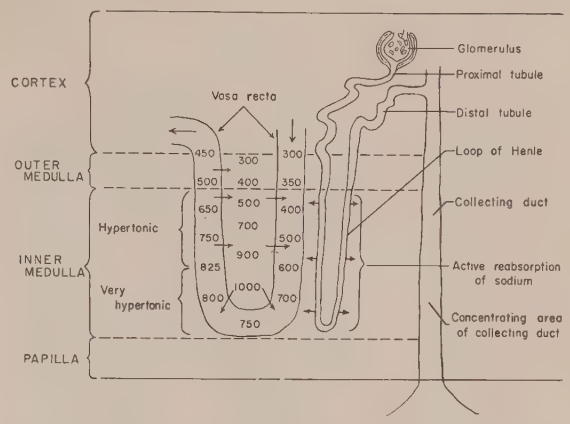


Figure 66. Counter-current mechanism by which the kidneys concentrate the urine.

fluid in the medulla has a very high concentration of sodium chloride. As the tubular fluid passes through the loop of Henle large quantities of electrolytes are usually actively reabsorbed by the tubular epithelium, so that the tubular fluid is dilute when it empties into the distal tubule. If the fluid passes without further change through the collecting tubule into the renal pelvis, the urine will be dilute. However, if large quantities of water should be reabsorbed from this fluid in the collecting tubule, the urine would become highly concentrated. Thus, if we should simply have a means for increasing or decreasing the amount of water reabsorbed from the collecting tubules, a mechanism would exist for changing the degree of concentration of the urine. Such a mechanism is the following:

The epithelial lining of the collecting tubule is almost completely impermeable to all the dissolved substances of the tubular fluid but has a variable permeability to water. A hormone released from the neurohypophysis, called *antidiuretic hormone,* causes the pores of the collecting tubules to become extremely permeable to water, but, on the other hand, when antidiuretic hormone is not present, the pores become almost entirely impermeable to water. Therefore, in the absence of antidiuretic hormone the dilute fluid from the distal tubules passes all the way through the collecting tubules into the renal pelvis, forming a dilute urine. On the other hand, *in the presence of antidiuretic hormone, most of the water in the tubular fluid is absorbed by osmosis* into the highly concentrated interstitial fluids of the medulla, so that the urine becomes highly concentrated. We can now see the importance of the counter-current mechanism which was responsible in the first place for developing a high concentration of sodium chloride in the medulla,

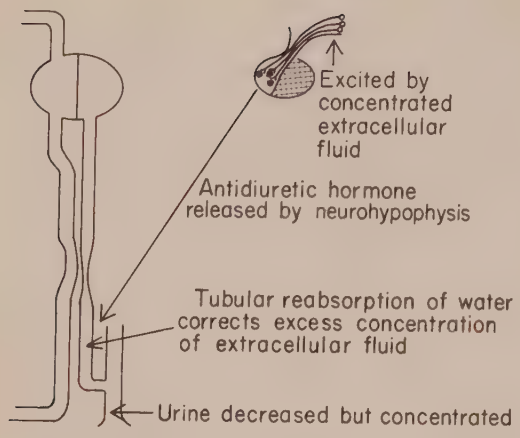


Figure 67. Osmoreceptor system for regulating crystalloidal osmotic pressure in the extracellular fluid.

thereby creating an osmotic gradient that can concentrate the urine.

In the absence of antidiuretic hormone the urine is diluted to a concentration about one-fourth that of the extracellular fluid, while in the presence of antidiuretic hormone, the urine is concentrated to about three times the concentration of the extracellular fluid.

Function of the osmoreceptor system in regulating the extracellular crystalloidal osmotic pressure. The complex renal mechanisms for diluting and concentrating the body fluids would not be of any importance if the body did not have some mechanism for increasing or decreasing the rate of secretion of antidiuretic hormone. Figure 67 illustrates the *osmoreceptor control system* that controls antidiuretic hormone secretion; this can be explained as follows: Located in the *supraoptic nuclei* and close thereby in the hypothalamus of the brain are a number of specialized neurons called *osmoreceptors.* These have small fluid chambers that swell when the body fluids become too dilute and contract when the body fluids become too concentrated. In the contracted state the osmoreceptors emit impulses that pass downward into the neurohypophysis and cause the release of *antidiuretic hormone.* The antidiuretic hormone then passes by way of the blood to the kidneys where it causes increased reabsorption of water from the collecting tubules.

Now, let us recapitulate the mechanism by which the crystalloidal osmotic pressure of the extracellular fluids is regulated. If the concentration of the fluids becomes too great, the osmoreceptors contract and emit impulses to the neurohypophysis. As a result, antidiuretic hormone is released which causes the kidneys to reabsorb excessive amounts of water while letting electrolytes continue to pass into the urine. As a result, electrolytes are lost while water is retained in the body fluids. Thus, the concentration of the body fluids becomes diluted, returning toward normal. Conversely, if the concentration becomes dilute, the osmoreceptors stop emitting impulses, the neurohypophysis stops releasing antidiuretic hormone, and the kidneys excrete a very dilute urine, thereby causing loss of water from the body fluids and concentration of the fluids back toward normal.

WATER DIURESIS. Figure 68 demonstrates the function of the osmoreceptor system. The curve depicts the rate of urine output per minute. At zero minutes, the subject drinks several glasses of water. In approximately 20 minutes the water has been absorbed into the extracellular fluid, diluting it. At this point the osmoreceptor system causes the kidneys to release excessive quantities of very dilute urine. This effect is called *water diuresis.* After an hour or more of losing the dilute urine, composed mainly of water, the concentration of the extracellular fluid returns to normal. Therefore, the urine flow and urine concentration also return to normal.

Regulation of Acid-Base Balance

The regulation of acid-base balance actually means the regulation of hydrogen ion concentration in the body fluids. When the hydrogen ion concentration is great, the fluids are *acidic;* when the hydrogen ion concentration is slight, the fluids are *basic.* The chemical reactions of the cells depend very greatly on the hydrogen ion concentration, which is the reason why the acid-base balance must be regulated very exactly.

The normal concentration of hydrogen ions in the body fluids is 4×10^{-8}. Usually this is expressed in terms of *pH,* which is the logarithm of the reciprocal of the hydrogen ion concentration. The pH of extracellular fluid and blood

$$\left(\text{the logarithm of } \frac{1}{4 \times 10^{-8}}\right)$$

is 7.4. A pH of less than 7.4 is on the acidic side, and a pH greater than 7.4 is on the basic side. The acid-base regulatory systems of the body are all geared toward maintaining a normal pH of approximately 7.4 in the extracellular fluid. Even in disease conditions it almost never becomes more acidic than 7.0 nor more basic than 7.8.

Regulation of acid-base balance by chemical buffers. All of the body fluids contain *acid-base buffers.* These are chemicals that can combine readily with any acid or base in such a way that

they keep the acid or base from changing the pH of the fluids greatly. One of the chemical systems that performs this function is the *bicarbonate buffer*, which is present in all body fluids. The bicarbonate buffer is a mixture of carbonic acid (H_2CO_3) and bicarbonate ion (HCO_3^-). When a strong acid is added to this mixture it combines immediately with the bicarbonate ion to form carbonic acid. Carbonic acid is an extremely weak acid; therefore, the buffer system changes the strong acid into a weak one and keeps the fluids from becoming strongly acid. On the other hand, when a strong base is added to this mixture, the base immediately combines with the carbonic acid to form water and a neutral bicarbonate salt. Loss of the weak acid and the addition of the neutral salt hardly affect the hydrogen ion concentration in the body fluids. Thus, it can be seen that this mixture of carbonic acid and bicarbonate ion protects the body fluids from becoming either too acidic or too basic.

Other important buffers are *phosphate* and *protein buffers*. These are especially important for maintaining normal hydrogen ion concentrations in the intracellular fluids, because their concentrations inside the cells are many times as great as the concentration of the bicarbonate buffer.

In essence, the buffers of the body fluids are the first line of defense against changes in hydrogen ion concentration, for any acid or base added to the fluids immediately reacts with these buffers to prevent marked changes in the acid-base balance.

Respiratory regulation of acid-base balance. Carbon dioxide is continually formed by all cells of the body as one of the end-products of metabolism. This carbon dioxide combines with water to form carbonic acid in accordance with the following reaction:

$$CO_2 + H_2O \leftrightarrows H_2CO_3$$

In other words, even the normal metabolic processes are always pouring acids into the body fluids. On the other hand, one of the functions of respiration is to excrete carbon dioxide through the lungs into the atmosphere. Normally, respiration removes carbon dioxide at the same rate it is being formed. If, however, pulmonary ventilation decreases below normal, carbon dioxide will not be excreted normally but instead will pile up in the body fluids, causing the concentration of carbonic acid to increase also. As a result, the hydrogen ion concentration rises. On the other hand, if the rate of pulmonary ventilation rises above normal, the opposite effect occurs; carbon dioxide is blown off at a more rapid rate than it is being formed, thereby decreasing the carbon dioxide and carbonic acid concentrations. Complete lack of breathing for a minute reduces the pH of the extracellular fluid from the normal of 7.4 down to about 7.1, while very active overbreathing can increase it to 7.7 in about one minute. Thus the acid-base balance of the body can be changed greatly by over- or under-ventilation of the lungs.

CONTROL OF RESPIRATION BY THE HYDROGEN ION CONCENTRATION OF THE BODY FLUID. In the preceding paragraph the effect on the acid-base balance of changing the rate of ventilation was discussed. In this paragraph the opposite effect of acid-base balance on respiration is discussed. A high hydrogen ion concentration stimulates the respiratory center in the medulla of the brain, greatly enhancing the rate of ventilation. Conversely, a low hydrogen ion concentration depresses the rate of ventilation. This effect of hydrogen ion concentration on the activity of the respiratory center affords an automatic mechanism for maintaining a fairly constant pH of the

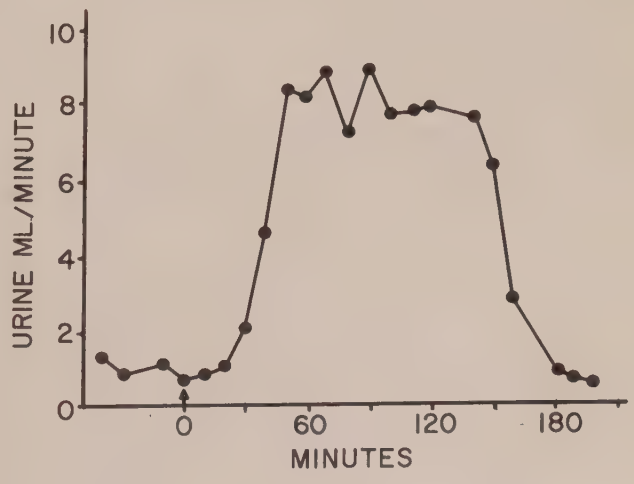


Figure 68. Water diuresis caused by the osmoreceptor system. (Redrawn from Smith: *The Kidney: Structure and Function in Health and Disease.* Oxford University Press.)

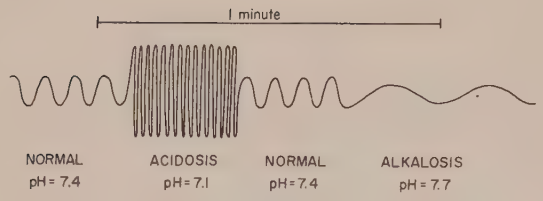


Figure 69. Effect of acidosis and alkalosis on respiration.

body fluids. That is, an increase in hydrogen ion concentration increases the rate of ventilation, which in turn removes carbonic acid from the fluids. Loss of the carbonic acid decreases the hydrogen ion concentration back toward normal. Conversely, diminished hydrogen ion concentration depresses the ventilation, and the hydrogen ion concentration rises back toward normal.

This respiratory mechanism for regulating acid-base balance reacts almost immediately when the extracellular fluids become either too acidic or too basic. Figure 69 illustrates the effect on the respiration of adding first a small amount of acid to the blood, and then later a small amount of alkali. The up and down curves represent the depth of respiration, and the rapidity of the curves represents the frequency of respiration. Note that acidosis greatly increases both the depth and rate of respiration, while alkalosis depresses respiratory function greatly, the depth of respiration becoming very slight and the rate very slow. This respiratory mechanism is so effective for regulating acid-base balance that it usually can return the pH of the body fluids almost to normal within a few minutes after an acid or alkali has been administered.

Renal regulation of acid-base balance. A number of other acids besides carbonic acid are continually being formed by the metabolic processes of the cells. These include large quantities if phosphoric acid and small quantities of sulfuric, uric, and keto acids, all of which, on entering the extracellular fluids, can cause acidosis. The kidneys normally rid the body of these excess acids as rapidly as they are formed, preventing an excessive build-up of hydrogen ions. These acids are generally called *metabolic acids,* in contradistinction to carbonic acid, which is called a *respiratory acid.*

In rare instances too many basic compounds enter the body fluids rather than too many acidic compounds. In general, this occurs when basic compounds are injected intravenously or when the person ingests a large quantity of alkaline foods or drugs.

MECHANISMS BY WHICH THE KIDNEYS REGULATE ACID-BASE BALANCE. The kidneys regulate acid-base balance by (1) excreting hydrogen ions into the urine when the extracellular fluids are too acidic and (2) excreting basic substances, particularly sodium bicarbonate, into the urine when the extracellular fluids become too alkaline.

Excretion of hydrogen ions and reabsorption of sodium. The distal tubular epithelium continually secretes hydrogen ions into the tubular fluid. These ions combine mainly with sodium salts in the tubular fluid to form weak acids and to free the sodium ion that is bound in the salt. The sodium in turn is absorbed through the tubular wall into the body fluids. Thus, there is a net exchange of hydrogen and sodium ions, the hydrogen ions passing into the urine and the sodium being reabsorbed. The precise manner in which this mechanism is used to regulate acid-base balance is quite complex, but it can be paraphrased in a few simple sentences as follows:

Ordinarily, the amount of hydrogen ions secreted into the tubules is almost exactly equal to the quantity of bicarbonate ions in the tubular fluid. The hydrogen ions can be considered to be an acidic substance and the bicarbonate ions an alkaline substance. The combination of the hydrogen ions and the bicarbonate ions forms carbonic acid which in turn splits into carbon dioxide and water. The carbon dioxide is reabsorbed from the tubules into the body fluids and then is excreted through the lungs. Thus, we have an acidic substance and a basic substance being destroyed, and the overall acid-base balance does not change when equal amounts of acidic and basic substance are destroyed. However, if the extracellular fluids become very acidic, the quantity of bicarbonate ions in the glomerular filtrate decreases. Therefore, now, far more hydrogen ions are secreted into the tubules than there are bicarbonate ions available to react with them. As a result, the hydrogen ions combine with the buffers of the tubular fluid and are carried into the urine. Thus, the acidic hydrogen ions are lost from the body fluids, returning the fluids to normal acid-base balance.

On the other hand, if the extracellular fluids become too alkaline, the quantity of bicarbonate ions in the glomerular filtrate becomes far greater than the quantity of hydrogen ions secreted by the tubules. The excess bicarbonate ions pass on into the urine combined mainly with sodium in the form of sodium bicarbonate. Therefore, a large amount of base is lost so that the body fluids become more acidic, returning the acid-base balance again back toward normal.

Ammonia secretion. Sometimes the body fluids become so acidic that it is impossible for the buffers in the tubules to carry all of the secreted hydrogen ions into the urine. When this happens, the distal tubules begin to secrete ammonia. The

ammonia combines with the hydrogen ions to form ammonium ions, which displaces sodium from the buffers and allows the sodium to be reabsorbed into the extracellular fluids. The ammonium ions then pass on into the urine combined in the form of a salt.

In summary, the renal mechanisms for regulating acid-base balance remove hydrogen ions from the extracellular fluids when the hydrogen ion concentration becomes too great, and they add hydrogen ions while removing sodium ions when the hydrogen concentration becomes too low. This principle is illustrated in Figure 70 which shows at point A a pH of about 7.5 in the extracellular fluid. Because this is on the alkaline side, the pH of the urine becomes very alkaline because of loss of alkaline substances from the body fluids. On the other hand, at point B the pH has fallen to 7.3, and the pH of the urine has become very acid because of loss of large quantities of acidic substances from the body fluids. In both of these instances the loss of alkaline or acidic substances returns the pH toward normal.

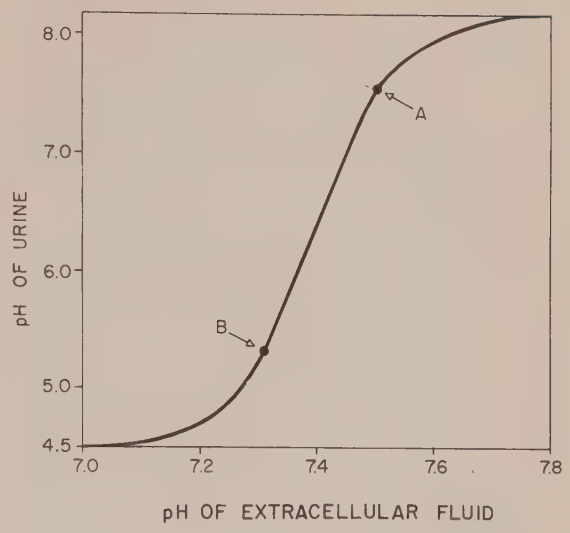


Figure 70. Effect of extracellular fluid pH on urine pH.

Abnormalities of acid-base balance. Many disorders of the respiratory system, the kidneys, or the metabolic systems for forming acids and bases can cause serious derangement of the acid-base balance. Some of the effects of these conditions are shown in Figure 71. The normal pH of the blood is 7.4, but intense over-ventilation can cause the pH to rise sometimes to as high as 7.8. On the other hand, asphyxia, which means extreme decrease in ventilation, causes a buildup of carbon dioxide and carbonic acid in the body fluids, thereby promoting acidosis, in which the pH of the blood sometimes falls to as low as 7.0.

A common cause of alkalosis is ingestion of alkaline drugs used for treatment of gastritis or stomach ulcers. The drugs sometimes are absorbed into the body fluids in quantities greater than the kidneys can remove, resulting in alkalosis.

Loss of large quantities of fluids from the intestinal tract at times causes acidosis but at other times alkalosis. For instance, vomiting large quantities of hydrochloric acid from the stomach gradually depletes the acid reserves of the extracellular fluid and causes alkalosis. On the other hand, vomiting of fluids from the lower intestine or loss of fluids as a result of diarrhea usually causes acidosis. This is because secretions from the lower intestine contain large quantities of sodium bicarbonate. When this basic salt is lost from the body fluids it is immediately replaced by carbonic acid formed by the reaction of carbon dioxide with water. Therefore, the net effect is loss of sodium ions and gain of hydrogen ions, causing acidosis.

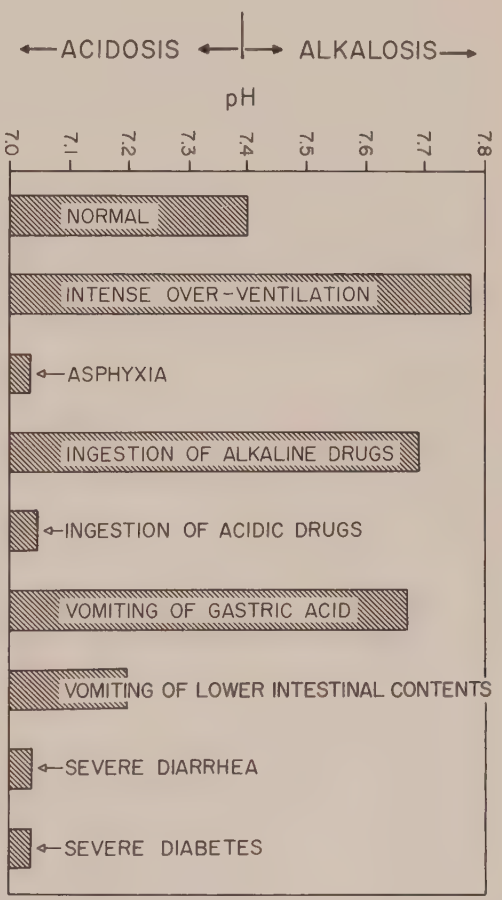


Figure 71. pH of the body fluids in various acid-base disorders.

Finally, Figure 71 shows extreme acidosis that sometimes results from severe diabetes. In diabetes many of the fats normally used by the body for energy are not completely metabolized but instead are broken into substances called keto acids that build up in the body fluids, producing extreme acidosis.

EFFECTS OF ACIDOSIS AND ALKALOSIS ON BODILY FUNCTIONS. Acidosis generally promotes depressed mental activity culminating in coma and death. Usually the afflicted person will pass into coma when the pH of his extracellular fluid falls below approximately 6.9.

On the other hand, alkalosis causes overexcitability of the nervous system, often resulting in excessive initiation of impulses. These may cause tetanic contraction of all the muscles and can actually kill because of convulsions or other derangements of nervous activity.

Regulation of Blood Volume

The normal blood volume is almost exactly 5000 ml., and it rarely rises or falls more than a few hundred milliliters from this value. Two principal mechanisms for maintaining this constancy are (1) the capillary fluid shift mechanism and (2) the kidney mechanism.

The capillary fluid shift mechanism. When the blood volume becomes too great, the pressures increase in all the vessels throughout the body, including the capillaries. The normal pressure in the capillaries is about 18 mm. Hg, as illustrated in Figure 72. When capillary pressures rise above this value, fluid automatically leaks into the tissue spaces as we explained in Chapter 6. Thus, the blood volume decreases back toward normal. When the capillary pressure has returned to normal, further loss of fluid into the tissue spaces ceases. Conversely, when the blood volume falls too low, the capillary pressure falls, and fluid is then absorbed from the interstitial spaces, increasing the blood volume again back toward normal.

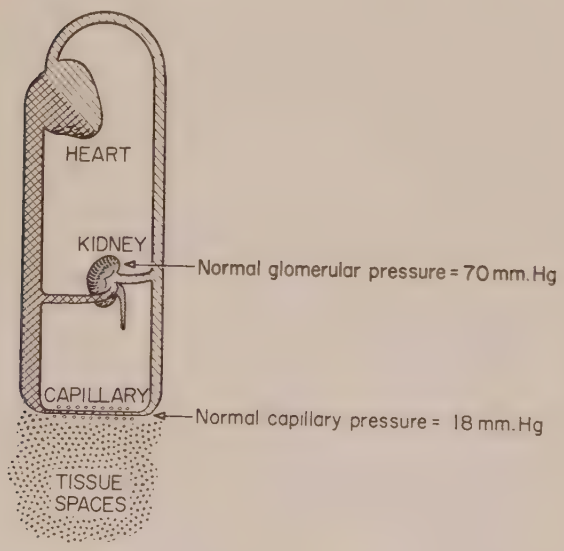


Figure 72. The capillary and kidney systems for regulating blood volume.

The kidney mechanism. As was explained in the previous chapter, when the glomerular pressure in the kidneys rises, the amount of glomerular filtrate and consequently the amount of urine formed by the kidneys greatly increase. Figure 72 also shows that the normal glomerular pressure is about 70 mm. Hg. An increased blood volume raises this to a higher level by two different mechanisms: First, the increased volume increases the arterial pressure, which causes increased flow of blood through the afferent arterioles into the glomeruli, thus raising their pressure. Second, the increased vascular pressure stretches the blood vessels in the thorax and neck that are supplied with stretch receptors called *pressoreceptors*. Stretching these initiates impulses that pass to the medulla of the brain. There they inhibit the normal sympathetic nerve impulses to the kidneys, which allows the renal afferent arterioles to dilate, thus increasing the blood flow into the kidney and correspondingly increasing the amount of urine formed.

One will immediately recognize this as a feedback mechanism by which the blood volume can be regulated; that is, an increase in blood volume initiates an increase in urinary output, which automatically decreases the blood volume back toward normal.

Regulation of Extracellular Fluid Volume

One can understand the regulation of extracellular fluid volume best by first recalling from Chapter 6 the mechanism by which negative interstitial fluid pressures develop throughout the body. It will be recalled that the colloid osmotic pressure of the blood is ordinarily considerably greater than the capillary pressure, which causes a tendency for fluid always to be absorbed from the interstitial spaces. This creates a negative pressure in the interstitial spaces of about −6 mm. Hg.

Therefore, in effect, the interstitial fluid volume is normally regulated to just that minimum volume that is necessary to fill the interstitial spaces. If ever the interstitial fluid volume becomes too great, the interstitial pressure rises markedly so

that fluid is absorbed into the blood far more easily than usual, which increases the blood volume. The increased blood volume in turn causes the kidneys to form increased amounts of urine. Thus, in effect, the interstitial fluid volume is lost into the urine by way of the blood.

Conversely, if the interstitial fluid volume falls too low, which decreases the interstitial fluid pressure below the normal value of −6 mm. Hg, fluid then leaks out of the blood to replenish the interstitial fluids. As a result, the blood volume falls too low, the rate of urine formation becomes greatly decreased, and increased amounts of water and salts are retained to build up the blood volume back to normal.

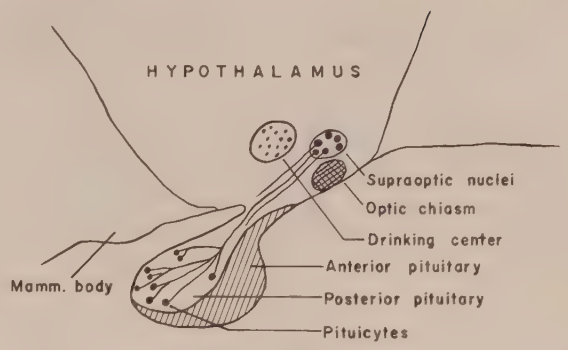


Figure 73. Location of the drinking center in the hypothalamus in relation to the supraoptic nuclei which control the antidiuretic hormonal system.

Regulation of Intracellular Fluid Volume

Each cell is autonomous in regulating its own fluid volume. It automatically transports those electrolytes from the extracellular fluid that are needed for intracellular function and these in turn pull water into the cell by creating osmotic pressure. Thus, because osmotic equilibrium between the intracellular and extracellular fluid compartments is maintained at all times, the fluid volume of each cell is directly proportional to the quantity of electrolytes pulled into the cell by its metabolic processes. Therefore, the regulation of intracellular fluid volume, unlike the regulation of the extracellular fluid volume, is not a direct function of the kidneys; yet, indirectly, the kidneys help in this regulation by maintaining appropriate constituents in the extracellular fluids.

Thirst, and Appetite For Electrolytes

In our discussion thus far, we have assumed that there is a constant intake of fluid and electrolytes and that regulation of body fluid constituents is carried out by increasing or decreasing the quantity of water or electrolytes lost into the urine. However, this is not entirely true, because a person's intake of both water and electrolytes is regulated at least partially by the thirst and appetite mechanisms. Both of these are controlled by special centers in the hypothalamus of the brain. We know much about the mechanisms that control thirst but unfortunately very little about those that control appetite for the electrolytes, except for the following well known facts: First, animals or human beings living far from regions where salt is plentiful have a definite craving for salt. Indeed, animals will seek out salt deposits called *salt licks* and will even fight for the salt just as they will fight for food. We know also that human beings, under special conditions, such as the pregnant mother, crave excessive amounts of salt and at times even other types of electrolytic substances such as sour pickles (acidic), or even calcium salts. Yet, the usual human being, living in our modern society of plenty, has such an abundance of the different types of salt in his diet that his appetite for a specific salt is almost never a significant factor in the regulation of the electrolyte content of his body fluids. On the other hand, the thirst mechanism does play an exceedingly important role in the hour by hour regulation of body fluid volumes and crystalloidal osmotic pressure.

The drinking center. Figure 73 illustrates a small area located in the hypothalamus a few millimeters behind the supraoptic nuclei, an area called the *drinking center.* When this center is stimulated electrically, the animal searches for the nearest water and begins to drink profusely. Also, injection of a concentrated solution of electrolytes into this center will likewise cause the animal to begin drinking.

One will note the close proximity of the drinking center to the osmoreceptors located in the supraoptic nuclei. When the body fluids become too concentrated, they excite both the drinking center and the osmoreceptors at the same time. The drinking center causes the animal to drink water, and the osmoreceptors cause the kidneys to conserve water.

REFERENCES

Adolph, E. F.: Early concepts of physiological regulations. *Physiol. Rev.* 41:737, 1961.

Adolph, E. F., Barker, J. K., and Hoy, P. A.: Multiple factors in thirst. *Am. J. Physiol.* 178:538, 1954.

Berliner, R. W., Levinsky, N. G., Davidson, D. G., and Eden, M.: Dilution and concentration of the urine and the action of antidiuretic hormone. *Am. J. Med.* 24:730, 1958.

Davenport, H. W.: *The ABC of Acid-Base Chemistry: The Elements of Physiological Blood-Gas Chemistry for Medical Students and Physicians.* 2nd Ed. Chicago, The University of Chicago Press, 1950.

Elkinton, J. R., and Danowski, T. S.: *The Body Fluids.* Baltimore, The Williams & Wilkins Company, 1955.

Farrell, G., and Taylor, A. N.: Neuroendocrine aspects of blood volume regulation. *Ann. Rev. Physiol.* 24:471, 1962.

Gottschalk, C. W., and Mylle, M.: Micropuncture study of the mammalian urinary concentrating mechanism: evidence for the countercurrent hypothesis. *Am. J. Physiol.* 196:927, 1959.

Langston, J. B., and Guyton, A. C.: Effect of epinephrine on the rate of urine formation. *Am. J. Physiol.* 192:131, 1958.

Schmidt-Nielsen, B., and Laws, D. F.: Invertebrate mechanisms for diluting and concentrating the urine. *Ann. Rev. Physiol.* 25:631, 1963.

Thorn, N. A.: Mammalian antidiuretic hormone. *Physiol. Rev.* 38:169, 1958.

Section Three

BLOOD AND IMMUNITY



THE BLOOD CELLS

9

Almost all the cells in the blood are *red blood cells,* and their major function is to transport oxygen from the lungs to the tissues. However, approximately one out of every 500 cells is a *white blood cell,* or *leukocyte,* which is called "white" because it is not colored by hemoglobin. The leukocytes can be divided into *granulocytes, monocytes,* and *lymphocytes,* and the granulocytes in turn can be divided into *neutrophils, eosinophils,* and *basophils.* White blood cells have several different functions, but the most important of these is to protect the body against invasion by disease organisms.

The blood also contains large numbers of *platelets,* which are often classified as white blood cells. However, the platelets are not really cells but instead very small fragments of a special type of white blood cell called the *megakaryocyte.* The platelets are essential for the clotting of blood, which will be explained in detail in Chapter 11.

THE RED BLOOD CELLS

Red blood cell count and hematocrit. The normal number of red blood cells in each cubic millimeter of blood is a little more than 5,000,000 in the male person and a little less than this in the female. The percentage of the blood made up of red blood cells is called the *hematocrit;* the normal figure is 45 per cent. The hematocrit is determined by centrifuging blood in a graduated tube of the type shown in Figure 74. This figure contrasts the hematocrit of normal blood, 45 per cent, with that of *anemic* blood, 15 per cent, and that of *polycythemic* blood, 65 per cent.

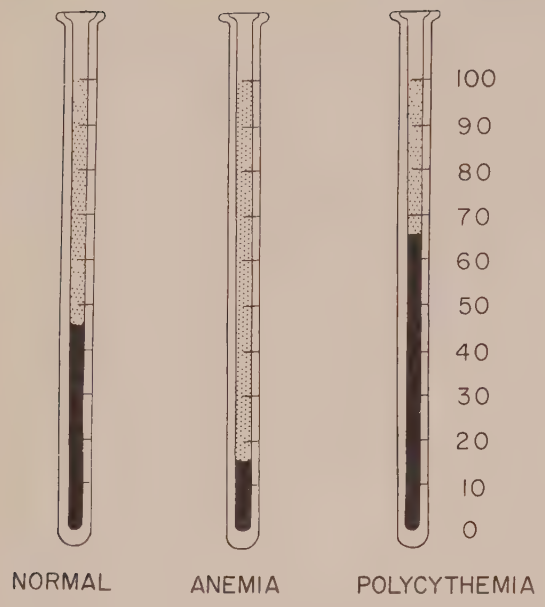


Figure 74. Determination of the hematocrit.

Organization of the red blood cell. The red blood cell is different from all other cells of the body in that it has already lost its nucleus so that it can no longer multiply after entering the blood stream.

The major constituents of the red blood cell are (1) a thin cellular membrane, and (2) a viscid solution of *hemoglobin*, having a concentration of about 33 per cent, on the inside. In addition to hemoglobin, many other substances, some of which are listed in Figure 75, are present in the red blood cell in small quantities. Some of these are proteins and lipids that make up the physical structures of the cell such as the cellular membrane and the internal framework. Others are enzymes, minerals, and less important compounds that are important for metabolism inside the cell as long as it remains in the circulation. Metabolism keeps the cell functional even though it is no longer growing or multiplying.

Transport of Oxygen by the Red Blood Cells

The principal function of the red blood cell is to carry oxygen from the lungs to the tissues. Oxygen diffuses through the pulmonary membrane into the blood, then through the cell membrane to combine with the hemoglobin inside the red blood cell. When the cell reaches a tissue capillary, the opposite effects take place: The hemoglobin releases the oxygen which then diffuses out of the capillaries into the tissues. This transport of oxygen from lungs to the tissues will be discussed in detail in connection with the subject of respiration in Chapter 17.

In some lower animals, hemoglobin circulates in the blood like plasma protein rather than being transported inside red blood cells. In the human being, however, the hemoglobin molecule is

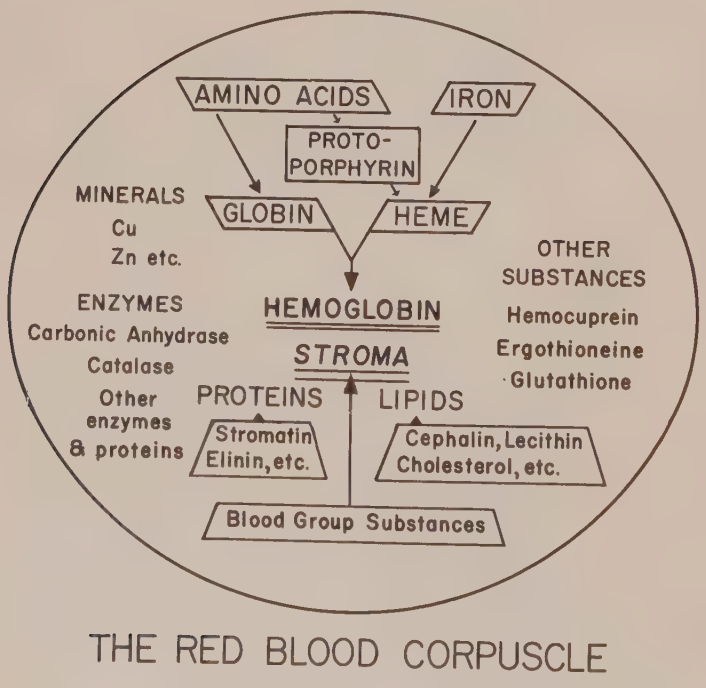


Figure 75. Constituents of the red blood cell. (Redrawn from Wintrobe: *Clinical Hematology*. Lea & Febiger.)

small enough that approximately 3 per cent of it diffuses through the capillary pores each time the blood passes through the systemic circulation. Therefore, the function of the red blood cell membrane is to keep the hemoglobin in the circulation, for as long as hemoglobin remains inside the cell, it cannot diffuse through the capillary membranes.

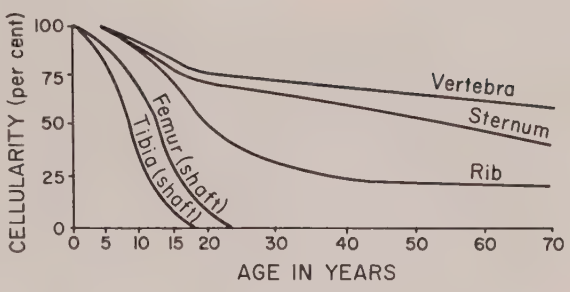


Figure 76. Rate of red blood cell production in the marrow of several different bones. (Redrawn from Wintrobe: *Clinical Hematology.* Lea & Febiger.)

Genesis of the Red Blood Cells

In the fetus red blood cells are formed in (1) the yolk sac of the embryo, (2) the liver, (3) the spleen, and (4) the bone marrow, but by the time the infant is born, red cell formation no longer occurs in any area except the bone marrow. Furthermore, the marrow of the membranous bones—that is, the skull, the vertebrae, the ribs, the ilia, and the sternum—forms more cells than the marrow of the long bones. As illustrated in Figure 76, red blood cells continue to be formed in the membranous bones throughout life, whereas they are formed in the shafts of the long bones only during preadolescence. It is evident from this figure, too, that the rate of red blood cell formation in all bones decreases with age. As a result, the number of red blood cells in the circulation also decreases, usually causing mild degrees of anemia in old age.

Formation of the cell structure. Figure 77 illustrates, to the left, the stages of red blood cell development, beginning with the *hemocytoblast* and ending with the *erythrocyte,* which is the mature red blood cell. In the final stage of development—that is, during the change from the *normoblast* to the erythrocyte—the nucleus disintegrates and is absorbed. During this absorption, which continues for a few hours to a few days, some of the breakdown products of the nucleus form a reticulum, or network, in the cell, and the cell at this time is called a *reticulocyte.* Finally, however, the last remnants of the nucleus are absorbed, leaving the cell essentially a bag of hemoglobin.

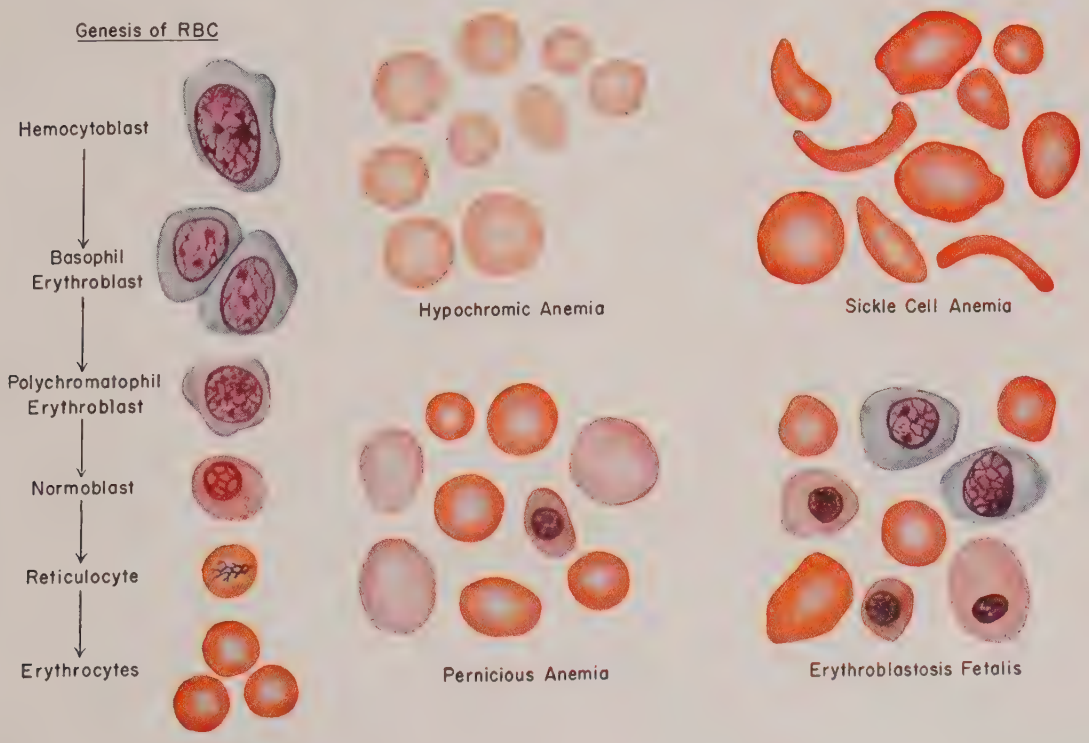


Figure 77. Genesis of red blood cells, and the blood pictures of several types of anemia.

The erythrocytes, on becoming mature, squeeze through the capillary walls into the marrow vessels, in this way entering the blood stream. Occasionally, when large numbers of red blood cells are being formed very rapidly, reticulocytes and normoblasts also enter the circulation before they have turned into erythrocytes. Indeed, the rapidity of red blood cell formation can be estimated simply by counting the number of reticulocytes and normoblasts in the blood.

SPECIAL SUBSTANCES NEEDED FOR RED CELL FORMATION. Special nutrients are necessary for the formation of the cell structure. These include especially *vitamin* B_{12} and *folic acid.* Without these the erythroblasts fail to proliferate rapidly but instead have a tendency to grow larger and larger without dividing as many times as usual. As a result, an oversized erythroblast, called a *megaloblast,* is formed. When the erythrocyte is finally formed it too is greatly oversized, and its membrane is very fragile.

Formation of hemoglobin. Synthesis of hemoglobin begins in the erythroblast and continues into the erythrocyte stage; even after young erythrocytes leave the bone marrow and pass into the blood stream, they continue to form hemoglobin for many days. The mechanism of synthesis is not completely known, but the five major stages are illustrated in Figure 78. The erythroblast first converts *acetic acid* into *ketoglutaric acid* which then undergoes a series of combinations and changes until the final hemoglobin molecule is formed.

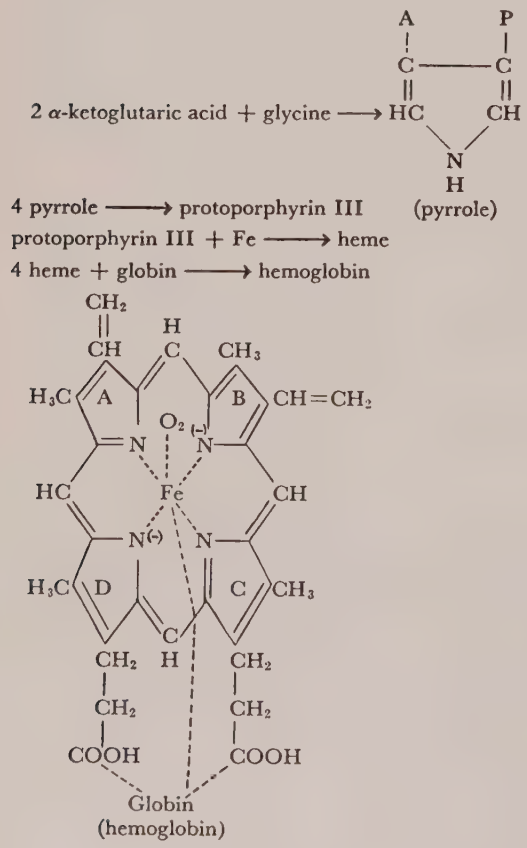


Figure 78. Formation of hemoglobin.

The iron portion (Fe) of the hemoglobin molecule is exceedingly important, for oxygen binds with this, as is discussed below. Experiments in animals have shown that *copper, cobalt, nickel,* and the vitamin *pyridoxine* are all also necessary for formation of hemoglobin, for without any one of these the quantity of hemoglobin in the red blood cell will be deficient. Yet the precise manner in which each of these helps in the formation of hemoglobin is unknown.

The most important feature of the hemoglobin molecule is its ability to combine loosely and reversibly with oxygen. The two positive valences of the ferrous iron are firmly bound with other portions of the hemoglobin molecule, but the oxygen binds with two of the six so-called "coordination" valences of the iron atom. These valencies form an extremely loose bond with the oxygen so that the combination is easily reversible. That is, in the lungs oxygen binds easily with the iron but separates from the iron equally as easily in the tissue capillaries to supply the tissue cells with oxygen.

IRON METABOLISM. The total quantity of iron in the entire body is about 4.5 grams, and approximately two thirds of this is in the hemoglobin of the blood. When the diet is deficient in iron, the concentration of hemoglobin in the red blood cells falls markedly, sometimes making it difficult to transport adequate amounts of oxygen to the tissues. Therefore, it is important to understand the relationship of iron metabolism to the formation of hemoglobin.

The amount of iron needed each day by the human being is 1 to 3 mg. About 1 mg. is excreted each day into the bowel and urine, and any intercurrent bleeding can account for loss of another milligram or two of iron from the body. For instance, menstrual bleeding each month in the female accounts for an average loss of about 1 mg. per day.

Only *ferrous* iron can be absorbed from the intestinal tract. Fortunately, however, the acid contents of the stomach can convert most inorganic ferric compounds into ferrous compounds prior to entry into the small intestine.

The first stage in absorption of the iron is the following: The ferrous iron is actively transported into the lining cells of the intestinal mucosa where it combines with a large protein called *apoferritin*

to form *ferritin.* Gradually small quantities of iron are released from the ferritin into the blood where it combines with a globulin to form another compound, *transferrin.* In this form, iron is carried throughout the body to be used where it is needed, for instance to the bone marrow for use in the formation of hemoglobin.

Once the body has all the iron that it can use, the plasma globulin which combines with iron to form transferrin becomes saturated. Now, no iron can be removed from the ferritin of the gastrointestinal mucosa so that all the apoferritin of the intestinal wall becomes converted to ferritin and therefore can no longer combine with iron from the intestine. Under these circumstances no additional iron is absorbed. Conversely, when the body is deficient in iron, the quantity of transferrin in the blood and the quantity of ferritin in the intestinal mucosa both diminish markedly, in which case, large amounts of apoferritin again become available to combine with ferrous iron from the intestine. Therefore, iron absorption again proceeds rapidly. In this way the absorption of iron from the intestine is automatically regulated so that the proper quantity of iron will be available at all times.

A large amount of iron is stored in the liver for use in periods of dietary deficiency. This iron is combined with *liver apoferritin* to form ferritin. When great excesses of iron are available to the body all of the apoferritin of the liver becomes ferritin. On the other hand, in times of iron deficiency the iron is released from the ferritin and is carried throughout the body in the form of transferrin to be used where it is needed.

In addition to the major need for iron in the formation of hemoglobin, it is also a key constituent of many other cellular oxidative metabolic substances such as myoglobin and the different cytochromes.

Life of the Red Blood Cell

Once the red blood cell has entered the circulation it circulates an average of 120 days before disintegrating. This long life of the red cell is evident from Figure 79, which shows the effect on the red blood cell count of cessation of red blood cell formation. The red blood cell count begins to fall rapidly because some of the older cells are already about to disintegrate. By 120 days almost all of the cells are gone, though a few very hardy cells, which had been released into the blood immediately before cessation of red blood cell production, may still continue to circulate for even a few more days.

During the life of the red blood cell some

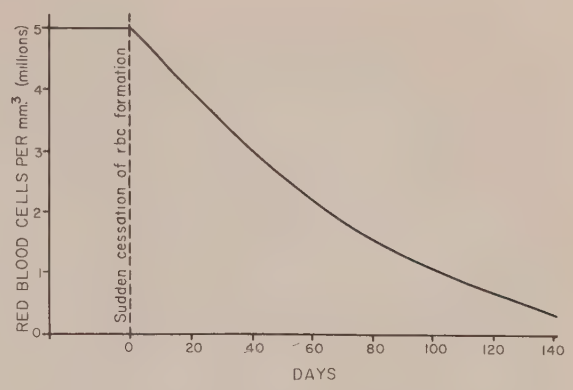


Figure 79. Decrease in red blood cell count following cessation of red blood cell formation.

metabolic processes continue to occur. The cytoplasm of the cell contains small amounts of oxidative and other enzymes which liberate the energy needed for maintaining integrity of the cellular structures. Even so, the cell eventually becomes fragile, and when it tries to pass through a tight capillary it finally breaks wide open, emptying its hemoglobin into the circulation. The fragments of the cell structure are engulfed by *reticuloendothelial* cells that line the walls of the capillaries in some organs such as the liver and spleen. These digest the fragments and release the breakdown products into the circulation in a dissolved state.

Fate of the hemoglobin. The hemoglobin released by the rupturing cells at first flows freely in the blood, but gradually most of it leaks into the tissue spaces. There it is engulfed by tissue reticuloendothelial cells and is digested by the intracellular enzymes. The breakdown products are *iron* and *bilirubin.* The bilirubin is excreted through the liver into the bowel as will be discussed in more detail in Chapter 29. The iron, on the other hand, is transported in the form of transferrin either to the bone marrow for formation of more hemoglobin or to the liver for storage in the form of ferritin.

Regulation of Red Blood Cell Production

Tissue oxygenation as the basic regulator of red cell production. Red blood cell formation is regulated principally by the need of the tissues for oxygen. When a person is exposed to an atmosphere low in oxygen, the rate of red blood cell production increases markedly, or when a person performs very severe and continuous exercise, the relative need for increased quantities of oxygen also increases the rate of red blood cell production. Thus, a well trained athlete fre-

quently has a red blood cell count of 6.5 million per cubic mm. in contrast to the normal of 5 million per cubic mm., and persons living at very high altitudes, such as some Peruvian and Chilean natives, often have red cell counts of 7.5 to 8 million. On the other hand, a lazy or bedridden person, whose tissues require little oxygen, may have a red cell count as low as 3 million.

The erythropoietic stimulating factor. Even though it is the need of the tissues for oxygen that causes increased red blood cell production, oxygen deficiency does not directly stimulate the bone marrow. Instead, it causes a humoral factor called *erythropoietic stimulating factor* (also known as *erythropoietin* or *hemopoietin*) to be formed, mainly in the kidneys. This substance in turn passes by way of the blood to the bone marrow and stimulates all phases of red blood cell production including the rate of cell maturation as well as the rate of erothyrocyte release from the bone marrow into the blood.

A deficiency of red blood cells in the circulation greatly enhances, and an excess of red blood cells greatly diminishes, the formation of erythropoietic stimulating factor; this in turn controls the rate of red cell production and returns the total quantity of red cells to normal. Thus, the number of red blood cells in the circulation is automatically regulated in response to the need of the tissues for oxygen, the greater the need, the greater the number of cells, and the less the need, the less the number.

Anemias

Anemia means a diminished concentration of red cells in the blood. Perhaps the most common cause of mild anemia is simply lack of exercise, for, as explained above, exercise is a stimulus to the production of red blood cells. A person with anemia from this cause can be treated with all the substances known to be necessary for red blood cell formation, including iron, vitamin B_{12}, folic acid, and proteins, and still he will continue to have the mild anemia.

In addition to this functional type of anemia, several disease conditions can also cause very serious anemias. Some of these are the following:

Blood loss anemia. Either acute blood loss or chronic blood loss can cause anemia. Acute loss of blood will immediately decrease the total blood volume, but within approximately 24 hours the plasma portion of the blood will have been replenished by ingested fluids or by fluids derived from the tissues. The red blood cells, however, require several weeks to be replenished. Consequently, after blood loss the red cells are diluted and their concentration in the blood is decreased.

In chronic blood loss the bone marrow becomes hyperactive, attempting to replace the lost red blood cells as rapidly as possible, but frequently it cannot keep up with the rate of loss. When blood is lost, some of the body's store of iron is lost as well, for approximately two thirds of all the iron in the body is present in the hemoglobin. Chronic blood loss is likely to deplete the body of iron, leaving insufficient quantities to form new hemoglobin. Consequently, in chronic blood loss anemia, the concentration of cells is reduced, and the quantity of hemoglobin in the cells that are formed is also far less than normal.

Iron deficiency anemia. Even without chronic blood loss, a person who does not ingest a sufficient quantity of iron may develop anemia. The actual number of red blood cells may be almost normal, but their size will be small because of insufficient hemoglobin, and the concentration of hemoglobin in each cell may be greatly decreased. This gives rise to the condition called *hypochromic* anemia, the blood cells of which are illustrated in Figure 77. In this condition, even though a large number of cells may be circulating in the blood, the amount of oxygen that can be transported to the tissues is far less than normal.

Anemias caused by fragile cells. Anemia can also result from *fragile cells,* whose membranes rupture on the slightest provocation. The hemoglobin then becomes dissolved in the plasma, and the reticuloendothelial cells ingest the cellular membranes. One of the causes of fragile cells is *familial microcytosis,* a hereditary disease that causes the red blood cells to be much smaller than usual and spherical in shape, rather than having the normal flat baglike shape. When these tightly filled cells are compressed even slightly, such as when they attempt to squeeze through tight capillaries, they rupture.

Another type of anemia caused by destruction of the red blood cells occurs in *erythroblastosis fetalis.* In this condition a pregnant mother becomes immunized against her child's blood, and antibodies (a special type of protein) from the mother's body pass through the placenta into the baby to cause destruction of the child's red blood cells. As the anemia becomes severe in the infant, the production of red blood cells becomes extremely rapid in an attempt to compensate for the anemia, causing many very early cells, normoblasts and even erythroblasts, to be extruded from the bone marrow into the blood. Figure 77 shows the resulting blood picture.

Anemia caused by poor production of cells. A final cause of anemia is poor production of the red blood cells themselves. In the preceding discussion of the formation of red blood cells, it

was pointed out that certain nutrients are required for formation of red blood cells. Lack of any of these nutrients will result in anemia. When vitamin B_{12} or folic acid are deficient, the number of cells is greatly reduced, and those that are produced are large, odd shaped, and quite fragile. This type of anemia is known as *pernicious anemia*, the blood cells of which are shown in Figure 77. Figure 80 shows the effect of treating pernicious anemia with liver extract, which contains vitamin B_{12} and folic acid. Immediately after treatment, the number of red blood cells begins to increase, and the reticulocytes, an early form of red blood cell as explained above, become a very large per cent of the total number of cells in the circulation. In addition, the average cell volume, as indicated by the top curve, becomes progressively less as the macrocytes characteristic of pernicious anemia are replaced by normal cells.

On rare occasions the bone marrow completely stops producing red blood cells, in which case the person is said to have *aplastic anemia*. This often occurs very suddenly when the bone marrow is poisoned by some drug, or when the bone marrow is exposed to intense ionizing radiations such as x-rays or the gamma rays emitted by the explosion of an atom bomb. In aplastic anemia those cells that are already in the circulation are completely normal, but because the aging cells are no longer replaced by new cells, the red cell count diminishes during the next few weeks, giving a very severe anemia after about a month.

Physiologic effects of anemia. Anemia affects the functions of the circulation in two ways. First, the decrease in oxygen-carrying capacity of the blood decreases the rate of oxygen transport to the tissues, which results in poor tissue oxygenation and can cause extensive damage throughout the body or even death. The second effect is decreased viscosity of the blood, for the viscosity is determined almost entirely by the concentration of red blood cells. The resistance to blood flow decreases when the viscosity decreases. Therefore, the low viscosity allows very rapid flow of blood through the peripheral circulation, promoting excessive cardiac output, dilatation of the heart, and overworking of the heart. Consequently, one of the results of anemia is often heart failure because of the extra load imposed.

Polycythemia

Polycythemia means an excess of red blood cells. Mild polycythemia is normal in persons who exercise excessively and also in persons who live at high altitudes, but still other persons develop a condition called *polycythemia vera*, in which the rate of red blood cell production becomes far greater than normal even though there is no apparent physiologic basis for the increased pro-

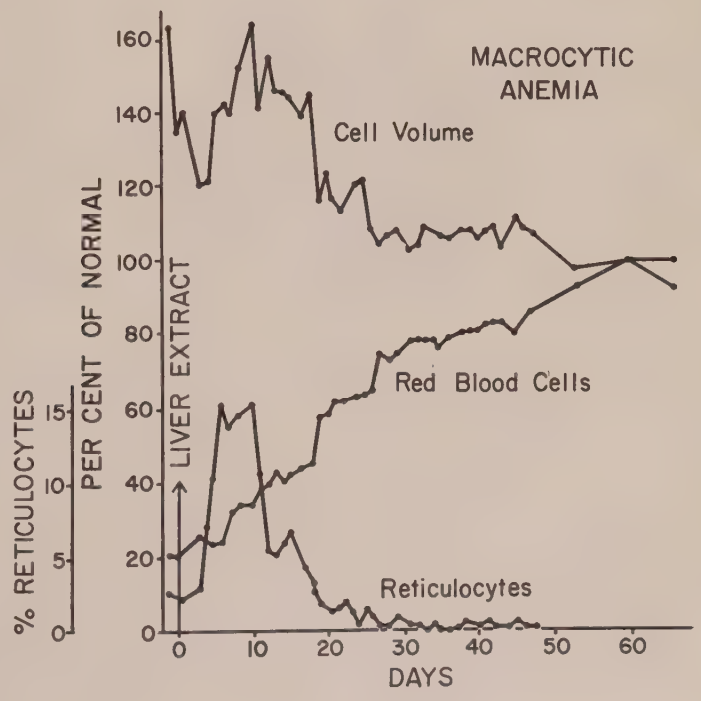


Figure 80. Effects of liver extract on the blood picture of pernicious anemia. (Redrawn from Wintrobe: *Clinical Hematology*. Lea & Febiger.)

duction. It is believed that this disease results from a tumorous condition of the bone marrow. Presumably the hemocytoblasts simply begin producing far too many red blood cells, and the normal regulatory mechanisms fail to stop this production. As a result, tremendous numbers of cells are poured into the circulation, maintaining a far greater red blood cell count than is desirable.

Physiologic effects. All of the important physiologic effects of polycythemia are caused by the increased viscosity of the blood that results. Sometimes the viscosity becomes as much as five times normal, which obviously greatly impedes the flow of blood through the circulatory system. Often the cells conglomerate in the small vessels, stopping flow entirely and resulting in many small blocks in the circulation. In approximately one third of polycythemic persons, blood flows through the systemic circulation so poorly that the arterial pressure rises far above normal, thereby placing an increased load on the heart. In this way, polycythemia like anemia—but for different reasons—can overload the heart and can also result in heart failure.

THE WHITE BLOOD CELLS

The total number of white blood cells in the circulation averages only 1/500 the number of red blood cells. Yet these cells are so important in protecting the body against disease that each type of white blood cell deserves special consideration.

Genesis of White Blood Cells

The larger proportion of the white blood cells are formed in the bone marrow along with the red blood cells. These include the neutrophils, the eosinophils, the basophils, and the platelets. All of these originate from the *myeloblast,* which, throughout the life of the person, continues to divide and to evolve into the different types of white cells. The stages of white cell development are shown in Figure 81. When each cell reaches maturity it is emptied from the bone marrow into the circulation.

Platelets as such are not produced in the bone marrow, but instead the bone marrow produces very large fragile cells called *megakaryocytes,* illustrated by cell number 3 in Figure 81. When these are mature they suddenly fragment into many minute parts which become the platelets. The platelets have a diameter approximately one-seventh that of a white blood cell and a volume about 1/300. Though the platelets are extremely important for the coagulation of blood, they have no specific relationship to the other white blood cells. They will be discussed in Chapter 11 in relation to coagulation, instead of in the present chapter.

Lymphocytes and monocytes are produced in lymphoid tissue throughout the body instead of in the bone marrow. Most of the lymphoid tissue is in the lymph nodes, which also filter the lymph as it flows through the lymphatics, but large amounts of lymphoid tissue are present in the spleen, thymus gland, and submucosa of the gastrointestinal and respiratory tracts. After lymphocytes are formed in the lymphoid tissue, they are emptied into the lymphatic channels and eventually flow along with the lymph into the circulation.

Life of the White Blood Cells

On studying the bone marrow with a microscope, it appears that the bone marrow produces approximately as many white blood cells as red blood cells. This makes it difficult at first to understand why the number of white blood cells in the circulation is very small in comparison with the number of red blood cells. However, the white blood cells, as opposed to the red cells, are capable of leaving the circulation by squeezing through the small pores of the capillaries and then passing through the tissue spaces. Furthermore, white blood cells rarely exist in the circulation or in the tissues for longer than a few days, for as they perform their functions to protect the body against disease, they themselves are usually destroyed. This explains why the number of white blood cells in the circulation is very slight.

Neutrophils

The neutrophils are the most important white blood cells for protecting the body against invasion by bacteria. These cells are approximately 12 microns in diameter and can pass rapidly through the pores of the capillaries by the process of *diapedesis* (squeezing through). On entering the tissue spaces they attack almost any agent that may be causing tissue damage.

Ameboid motion of the neutrophils. The neutrophils exhibit more ameboid motion than perhaps any other cell of the body. That is, they move through the tissues by (1) projecting a *pseudopodium*—a fingerlike extension—ahead of the main mass of the cell, (2) attaching this pseudopodium to a new position, and (3) pulling the remainder of the cell forward. This process

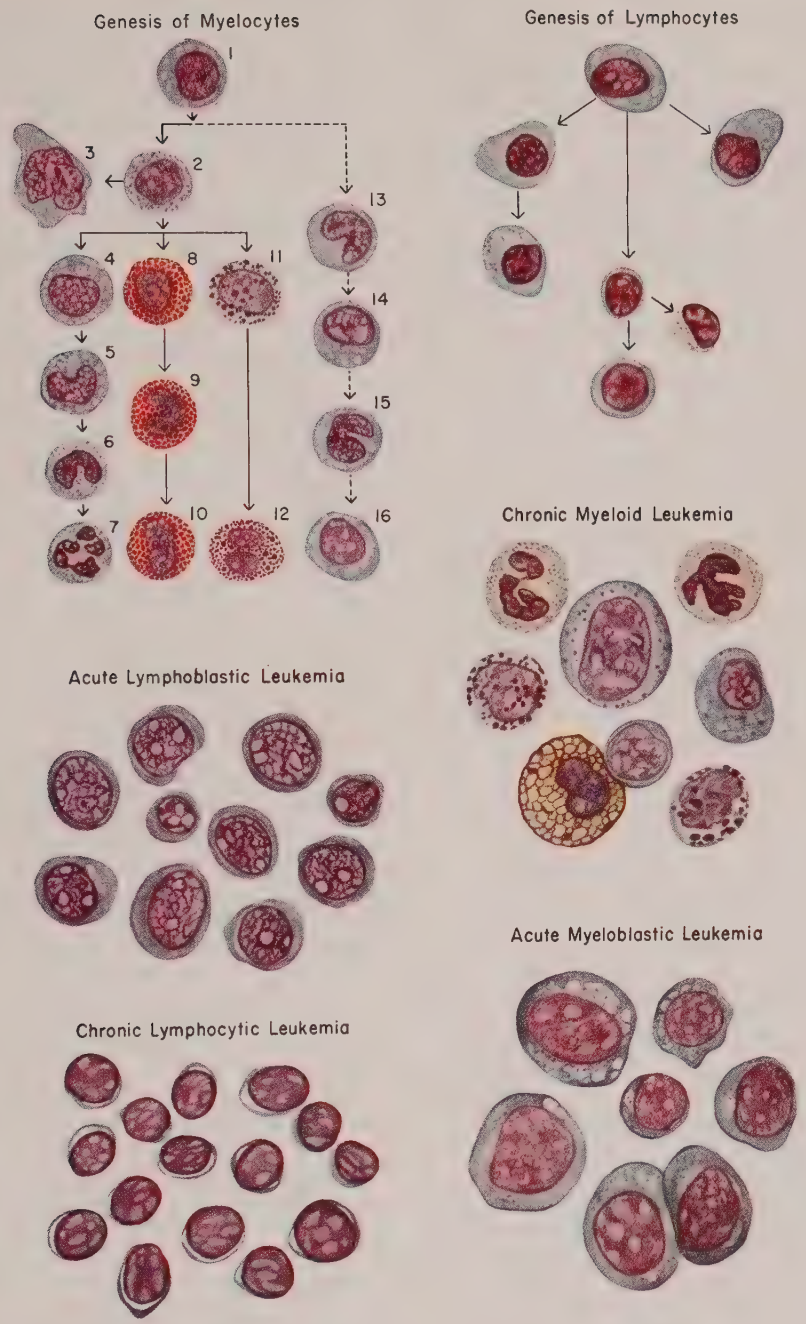


Figure 81. Genesis of the white blood cells, and the blood picture in several types of leukemia. Cell number 1 is a *myeloblast,* and number 3 is a *megakaryocyte.* Cells 4, 5, 6, and 7 illustrate the formation of *neutrophils;* 8, 9, and 10, the formation of *eosinophils;* 11 and 12, the formation of *basophils;* and 13, 14, 15, and 16, the formation of *monocytes.* (Redrawn in part from Piney: *A Clinical Atlas of Blood Diseases.* The Blakiston Co.)

is repeated over and over again until the cell reaches its destination.

The direction in which the pseudopodium projects and moves the cell is determined by chemical substances in the tissue spaces, a phenomenon called *chemotaxis* which may be explained as follows: When a tissue area is damaged it releases a substance, called *leukotaxine,* that diffuses in all directions. Obviously, the greatest concentration of leukotaxine will be in the damaged area, and its concentration will become progressively less farther and farther away from

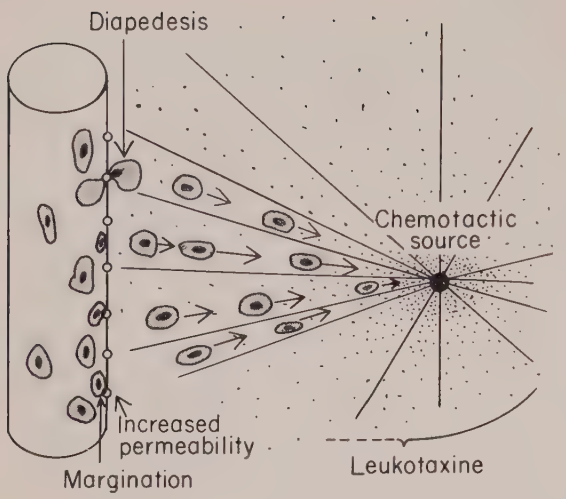


Figure 82. Movement of neutrophils by the process of *chemotaxis* toward an area of tissue damage.

the area of damage. The neutrophil, for reasons that are not completely understood, always moves toward an area of higher concentration of leukotaxine, until it finally reaches the damaged tissue.

Figure 82 exhibits the complete mechanism of the movement of neutrophils to an area of tissue damage, showing (1) increased porosity of the capillaries caused by toxins from damaged tissues, (2) adhesion of neutrophils to the wall of the capillary in the damaged area, a process called *margination*, (3) diapedesis of the neutrophils through the enlarged pores, and (4) ameboid movement of the neutrophils to the damaged area, guided by the process of chemotaxis.

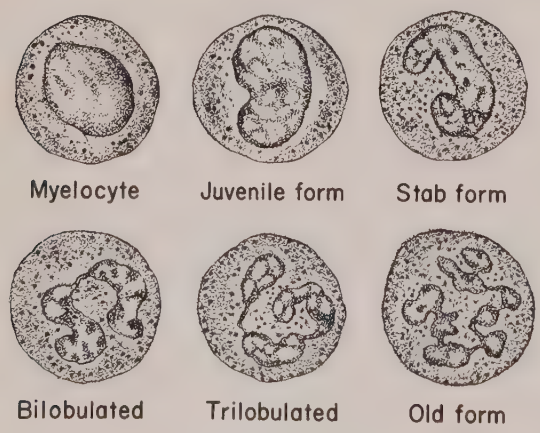


Figure 83. Neutrophils at different stages of aging, showing above the early forms, which appear in the circulation during rapid production of neutrophils, and below the older forms normally present.

Phagocytosis by the neutrophils. Neutrophils have the ability to ingest, or *phagocytize*, particles that are foreign to the tissues, including especially bacteria and to a lesser extent broken down tissue debris. When the neutrophil comes in contact with the particle, its membrane first attaches to the particle and then rapidly spreads in all directions around the particle so that the particle becomes engulfed in a deep well inside the cell.

Digestion of foreign particles by neutrophils. Neutrophils contain digestive enzymes that are capable of digesting most ingested particles. These enzymes are especially adapted to digest the proteins of bacterial bodies. After digestion, all that remains of a bacterium is many small chemical compounds which are readily dissolved in the intracellular fluids of the neutrophils.

Ordinarily, the neutrophil can phagocytize and digest 5 to 25 bacteria before the neutrophil itself becomes exhausted and dies, but on occasion the bacteria are so toxic to the neutrophils that only one or two may kill a neutrophil.

Neutrophilia. Neutrophilia, which means an increase in the number of neutrophils in the circulation, almost always occurs whenever tissues are damaged severely anywhere in the body. The damaged tissues, in addition to releasing leukotaxine which causes chemotaxis, also release another substance called *leukocytosis promoting factor*. This substance diffuses into the capillaries and flows in the blood to the bone marrow. Here it causes rapid release of leukocytes from the bone marrow into the blood. Because by far the largest number of leukocytes produced by the bone marrow is neutrophils, this substance increases especially the number of circulating neutrophils. Sometimes neutrophils are emptied into the circulation so rapidly that their number increases to two to three times normal in less than an hour.

The more rapid the release of leukocytes from the bone marrow, the earlier the form of the neutrophil. Figure 83 illustrates the early and old types of neutrophils, beginning with the myelocyte and progressing to the very old multilobulated form. Ordinarily neutrophils are not released into the circulation until they have progressed to the bilobulated or trilobulated stages, but under the stimulus of leukocytosis promoting factor, large numbers of stab, juvenile, and even myelocyte forms are also released.

When a much larger number of neutrophils than normal appears in the circulation, and especially when these include many early cells, one can be quite certain that somewhere in the body severe tissue damage is taking place. Most frequently this damage is caused by a serious bacterial infection, but other types of damage, such as that resulting from occlusion of a blood

vessel, damage to the tissues by trauma, or nuclear bomb radiation, can also lead to *neutrophilia.*

Monocytes

Monocytes function much like the neutrophils, though they have a considerably different appearance. They are often as large as 20 microns in diameter, they do not contain the granular substance characteristic of the neutrophils, and they do not have a multilobulated nucleus. Monocytes move through tissues by ameboid motion, but somewhat more slowly than do neutrophils. However, when monocytes finally reach the area of tissue damage, they are capable of engulfing several times as many bacteria, at times as many as 100 per cell, and far more tissue debris than can neutrophils. In other words, the phagocytic properties of monocytes are far superior to those of neutrophils, though the chemotactic and ameboid properties are much inferior.

Monocytes contain the same digestive enzymes as neutrophils plus others not present in the neutrophils, including especially lipases which can destroy the protective fatty shell of such bacteria as the tubercle bacillus.

When a bacterial disease lasts for weeks, the proportion of monocytes in the circulating white blood cells becomes greater and greater until finally their number may be as great as that of the neutrophils. Therefore, even though in the earlier stages of tissue damage almost all of the phagocytic cells reaching the damaged area are neutrophils, in the later stages the monocytes often outnumber the neutrophils. Thus, neutrophils are the most important white blood cells for resisting very acute infection, while monocytes are more important for resisting long-term, chronic infections. Also, because monocytes are large enough to phagocytize very large particles of tissue debris, they are much more important than neutrophils in cleaning up the tissues after an infection has been successfully combated.

Eosinophils

Eosinophils are closely related to neutrophils, though they have larger granules that stain strongly with acidic dyes. These cells exhibit chemotaxis, phagocytosis, and ameboid motion much the same as the neutrophils, but to so much less extent that they are not especially effective as scavengers in areas of tissue damage.

Unfortunately, the precise function of eosinophils is not yet known, but it has been suggested that they might detoxify foreign proteins, for they enter the blood in large numbers after foreign proteins are injected. Furthermore, many eosinophils are present in the mucosa of the intestinal tract and in the tissues of the lungs where foreign proteins normally enter the body.

The total number of eosinophils in the circulating blood increases greatly during allergic reactions. Here again, there is no reasonable theory to explain this effect, though it is possible that the allergic reaction releases toxic products from the tissues that might be detoxified by the eosinophils.

Another common cause of extremely large numbers of eosinophils in the blood is infection with parasites. For instance, in the condition known as *trichinosis,* which results from invasion of muscle by the trichinella parasite ("pork worm") after eating inadequately cooked pork, the percentage of eosinophils in the circulating blood occasionally rises to as high as 25 to 50 per cent of all the leukocytes. The cause of this is obscure, though it might again be a reaction in which the eosinophils are detoxifying foreign products produced by the parasites.

Basophils

The basophils are also similar to the neutrophils, except that the granules in their cytoplasm are even larger than those of the eosinophils and they stain darkly with basic dyes. The basophils exhibit very mild ameboid motion, phagocytic activity, and chemotaxis, but it is doubtful that these properties make the basophils of any significance in protecting the tissues against damage.

Though the precise function of the basophils is yet unknown, these cells are believed to be identical with the large *mast cells* which are found adjacent to the outside of the capillaries throughout the body. Mast cells are especially abundant in the pericapillary areas of the liver and the lungs, and they secrete a substance called *heparin.* Heparin diffuses into the blood to prevent blood coagulation, which is one of the reasons why the blood normally does not coagulate in the circulation. It is believed that the basophils circulating in the blood also release small quantities of heparin and thereby help to prevent intravascular blood coagulation.

Lymphocytes

Many of the lymphocytes in the blood are probably "spent" cells that have already performed their function in the lymph nodes before being released into the blood. Lymphocytes in the

lymph nodes help to detoxify substances flowing through the node in the lymph, and they produce small quantities of immune bodies, as described in the following chapter.

It has recently been learned that lymphocytes can become monocytes which can then penetrate tissues and perform a very valuable cleansing function in infected areas, as was described earlier in the chapter. Lymphocytes can also change into *tissue histiocytes, fibroblasts,* and *plasma cells* which are the cells that produce most of the immune bodies for protecting the body against bacteria and foreign toxins. Furthermore, recent experiments even indicate that lymphocytes can pass from the blood to the bone marrow and form hemocytoblasts and myeloblasts, consequently leading to the formation of red blood cells and granulocytes. If this is indeed true, then it is evident that lymphocytes are actually a very early form of cell not greatly different from the original germ cell from which the human being was formed. It is obvious, also, that lymphocytes have a multiple potential for transforming into other types of cells needed throughout the body.

Differential Count of White Blood Cells

A study of the number of different types of white blood cells in the circulation is called a "differential count." This count is made by accurately diluting a small amount of blood, placing this mixture in a calibrated glass counting chamber, and counting the different types of cells using a microscope. This can be of great value in diagnosing certain diseases. For instance, when the number of neutrophils is greatly increased, one has reason to believe that serious tissue damage is occurring, and that more likely than not the damage is being caused by bacterial infection. When the number of monocytes in the circulation is greatly increased, one suspects a chronic infection such as tuberculosis, chronic fallopian tube infection in the female, or some other long-term disease. A high eosinophil count can be of aid in diagnosing allergic conditions or parasitic infestation, and, finally, platelet counts are of special importance in determining the cause of excessive bleeding in certain patients.

Figure 84 illustrates the number of the different white blood cells in each cubic millimeter of blood at different ages. It will be noted that in the adult the total number of white blood cells is normally about 7500 per cu. mm. About 65 per cent of these are granulocytes, and of this 65 per cent approximately 62 per cent are neutrophils, 2 per cent eosinophils, and 1 per cent basophils. The normal proportion of lymphocytes in the white blood cells is approximately 30 per cent. The remaining 5 per cent are monocytes. In severe acute infections the percentage of granulocytes can rise to as high as 95 per cent, and the total number of white blood cells to as many as 30,000 to 50,000. In long-term chronic infections the percentage of monocytes can rise to as high as 40 per cent and the total number of white blood cells to 15,000 to 20,000.

Leukemia

Leukemia means an excess of white blood cells in the circulation caused by cancer of the white

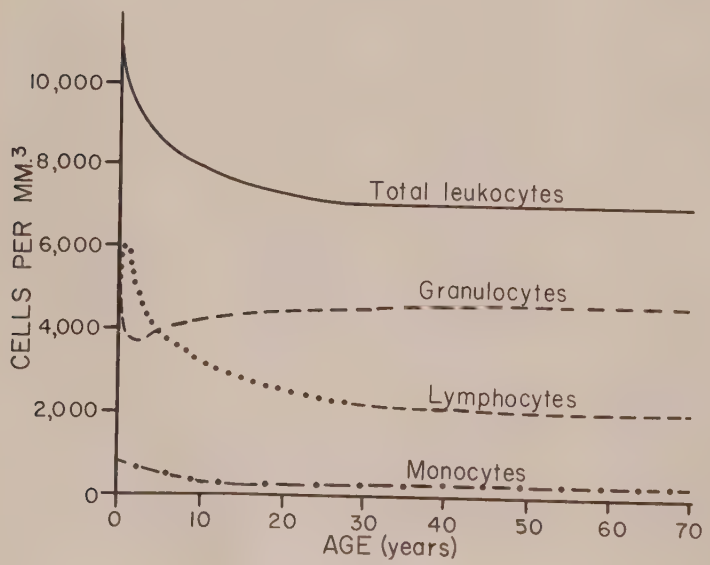


Figure 84. Concentration of the different types of leukocytes in the circulation at different ages.

cell producing tissue. The leukemias caused by cancer of the bone marrow are called *myelogenous leukemias,* while those caused by cancer of the lymphoid tissue are called *lymphocytic leukemias.* Leukemias have all degrees of severity. When the cancerous condition is spreading extremely rapidly, the cells released into the circulation are usually very early forms—in the case of myelogenous leukemia *myeloblasts,* and in the case of lymphocytic leukemia *lymphoblasts.* A person having one of these rapidly developing leukemias usually dies within six months to a year. The cause of death is the same as that in most rapidly developing cancers, namely, general debility of the body because of loss of nutrition to the cancerous cells.

In the slowly developing leukemias, the cells released into the circulation may be almost entirely normal and may have the appearance of neutrophils, eosinophils, basophils, monocytes, or small lymphocytes, depending on which particular type of cell in the bone marrow or lymph tissue has become cancerous. Often the total number of white blood cells will rise to as high as 100,000 to 200,000 per cu. mm. instead of the normal number of about 7500. Indeed, the number of circulating white blood cells is usually considerably greater in chronic leukemia than in acute leukemia, because in acute leukemia the cancerous condition frequently limits itself to the bone marrow or lymphoid tissue without releasing large numbers of cells into the circulation. Eventually, even the chronic type of leukemia will lead to inanition and death of the person, though frequently persons with this type of leukemia live as long as 12 to 20 years before this inevitable result.

The white cells found in the circulation in several different types of leukemia are illustrated in Figure 81. These show the large lymphoblasts and myeloblasts characteristic of the acute leukemias, and the small lymphocytes and relatively normal myelocytes characteristic of the more chronic leukemias.

Leukopenia

Leukopenia means decreased number of white blood cells in the circulation. This usually results from damage to the bone marrow by poisons, by toxic reaction to drugs, or by ionizing radiations from x-ray or nuclear bomb exposure. The lack of neutrophils, basophils, and eosinophils is called *agranulocytosis.* When the number of white blood cells falls extremely low the body becomes unprotected against bacterial invasion. Usually within three to four days very serious ulcerative infections begin to appear in the mouth, the colon, and other regions of the gastrointestinal or respiratory tracts, for these tracts always contain many pathogenic bacteria that normally are held in abeyance by continual action of the white blood cells. If the person is not treated with antibiotics or other bacteria-resisting drugs, he usually will die of fulminating infection within another day or two.

REFERENCES

Athens, J. W.: Blood: leukocytes. *Ann. Rev. Physiol.* 25:195, 1963.

Berlin, N. I., Waldmann, T. A., and Weissman, S. M.: Life span of a red blood cell. *Physiol. Rev.* 39:577, 1959.

Dacie, J. V.: *The Hemolytic Anemias.* 2nd Ed. New York, Grune & Stratton, Inc., 1958.

Dinning, J. S.: Nutritional requirements for blood cell formation in experimental animals. *Physiol. Rev.* 42:169, 1962.

Karnovsky, Manfred L.: Metabolic basis of phagocytic activity. *Physiol. Rev.* 42:143, 1962.

Linman, J. W.: *Factors Controlling Erythropoiesis.* Springfield, Ill., Charles C Thomas, 1960.

Low, R. N., and Freeman, J. A.: *Electron Microscopic Atlas of Normal and Leukemic Human Blood.* New York, McGraw-Hill Book Company, Inc., 1958.

Macfarlane, R. G., editor: *The Functions of the Blood.* New York, Academic Press, 1960.

Manwell, C.: Comparative physiology: blood pigments. *Ann. Rev. Physiol.* 22:191, 1960.

Rebuck, J. W., Bethell, R. H., and Monto, R. W.: *The Leukemias.* New York, Academic Press, 1957.

Weissman, S. M., Waldmann, T. A., and Berlin, N. I.: Quantitative measurement of erythropoiesis in the dog. *Am. J. Physiol.* 198:183, 1960.

Wilkinson, J. F.: *Modern Trends in Blood Diseases.* New York, Paul B. Hoeber, Inc., 1955.

THE RETICULOENDOTHELIAL SYSTEM, IMMUNITY, AND ALLERGY

10

This chapter presents the mechanisms by which the body protects itself against infections and toxins, including especially a discussion of the reticuloendothelial system and immunity. Also, we will discuss allergy, which is an unfortunate side effect of immunity.

THE RETICULOENDOTHELIAL SYSTEM

The reticuloendothelial system combats infection and toxins either by phagocytosis of foreign organisms or toxins or by formation of antibodies that destroy these. We have already discussed the white blood cells, which are considered by some physiologists to be part of the reticuloendothelial system. Other parts that aid in combating invading organisms include *Kupffer cells* that line the liver sinusoids; *reticulum cells* that line the sinusoids of the lymph nodes; phagocytic cells in the *spleen, bone marrow,* and *lymphoid tissue of the gastrointestinal tract;* and a special type of cells called *histiocytes* located throughout almost all tissues.

One will note that the reticuloendothelial cells are propitiously located for removal of invading organisms or toxins before they spread widely.

For instance, large numbers of colon bacilli are absorbed from the large intestine into the portal blood, but almost none of these succeed in passing through the liver sinusoids without being phagocytized by the Kupffer cells, which are illustrated in Figure 85. Indeed, motion pictures have shown that a colon bacillus, on coming in contact with the membrane of a Kupffer cell, enters the cell within the next 1/100 of a second, illustrating the extreme propensity of these cells for removal of foreign debris from the circulation.

The reticulum cells in the lymph nodes also play a special role in cleansing the body fluids, for all lymph must pass through the sinuses of lymph nodes before entering the blood. Therefore, fluids draining from an infected area, carrying with them live bacteria, are first cleansed by the lymph nodes prior to reaching the general circulation.

In virtually all tissues of the body are cells called *histiocytes* or *clasmatocytes* that have properties almost identical with those of the reticulum cells in the lymph nodes and the Kupffer cells in the liver sinuses. The tissue histiocytes are normally sessile cells awaiting the appearance of an organism to be phagocytized. However, when tissues are infected, histiocytes swell, become round, and become monocytes that can then move through the tissues to help in the cleansing process.

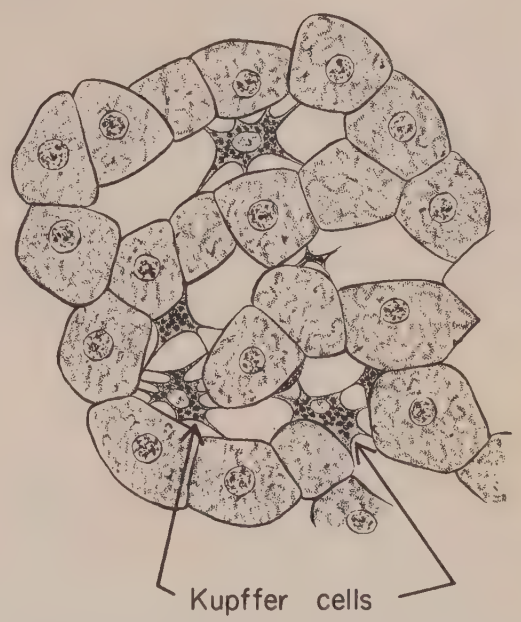


Figure 85. The Kupffer cells of the liver sinusoids. These are typical reticuloendothelial cells. (Redrawn from Smith and Copenhaver: *Bailey's Textbook of Histology,* The Williams & Wilkins Company.)

Inflammation, and the Walling Off of Infected Tissue Areas

Another important process for preventing the spread of infection is the process of *inflammation* which occurs in any tissue that becomes damaged. This process may be described as follows:

Damage to the tissues causes several substances to be released from the cells. One of these is *leukotaxine* and another is *necrosin.* Leukotaxine attracts polymorphonuclear neutrophils from the blood stream into the infected area as is described in the previous chapter. Necrosin has several important effects: First, it increases the permeability of the capillaries in the inflamed tissues, which allows fluid, proteins, and white blood cells to leak almost unabated into the area. Second, necrosin activates the fibrinogen in the leaked fluid, and thereby causes the fluid to clot in the same manner that blood clots, as is described in the following chapter. Clotting of these fluids prevents their flow from the damaged tissue area into the surrounding areas and, therefore, provides so-called "walling off" of the inflamed tissue from the surrounding tissues. This is especially important if the damage to the tissues is caused by bacteria, because it prevents the spread of bacteria.

IMMUNITY

Some bacteria and toxins are not phagocytized by the reticuloendothelial cells but instead must be destroyed in some other way. Fortunately, the body has still another line of defense called *immunity.* The blood and other body fluids contain special proteins called *antibodies* that react chemically with the invading organisms or toxins to destroy them.

Natural immunity. All human beings have natural immunity to certain diseases. For instance, distemper, a disease that kills approximately 50 per cent of all dogs that become infected with it, is totally non-infective to the human being, and the virus disease herpes simplex, which is lethal to rabbits, usually causes nothing more than fever blisters in the human being. This natural immunity results at least partly from the presence of *natural antibodies* in the plasma that have the ability to combine with the invading toxins or bacteria. However, as we shall see later in the chapter, most of the antibodies are not natural but instead are formed only *after a person has been exposed to the disease agent.*

Specific immunity. In addition to natural immunity, the human being can develop an extremely high degree of specific immunity against toxins or organisms to which he becomes exposed. For instance, a very minute quantity of tetanus toxin, approximately one millionth of a gram, is sufficient to cause death of a person. However, if a person is exposed several times to less than a lethal dose of this toxin, he can develop so many antibodies to the toxin that as much as 100,000 times the original lethal dose can be injected without causing death.

The Immune Process

Antigens. When bacteria or viruses enter the body, some of them eventually die, then disintegrate, and their proteins are emptied into the circulating fluids. These proteins, as well as the proteins of the various toxins, are called *antigens*, and they initiate the immune process which causes the production of antibodies as is described in the following paragraphs.

Formation of antibodies. Antibodies are formed to a slight extent in all cells, but certain cells are specialists in this task, the most important of which are the *plasma cells* in lymphoid tissue.

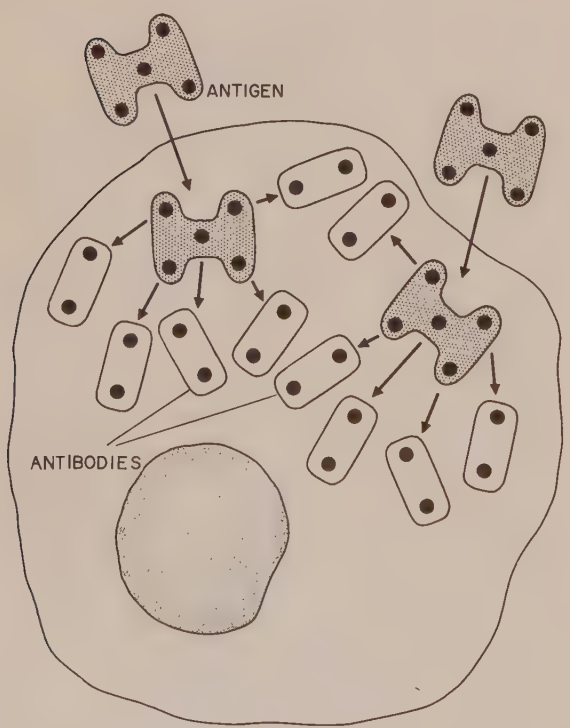


Figure 86. Formation of antibodies by a plasma cell.

In addition, the large lymphocytes and the reticulum cells lining the sinus walls of the bone marrow, spleen, and liver can form moderate quantities of antibodies.

The mechanism by which antibodies are formed is still conjectural. The first step is phagocytosis of the antigen by the plasma cell. Then the general process is probably that shown in Figure 86. The dumbbell-shaped object is an antigen, which is believed to alter some of the cellular enzymes in such a way that they thereafter produce antibodies, shown in the figure as the small, oblong rods. These are proteins that have mirror-image electrical charges on their surfaces, opposite in polarity to those on the antigens. Because of the opposite charges, the antibodies have an intense affinity for the antigen. After the cell has produced a large quantity of antibodies, the cell membrane ruptures and spills most of its cytoplasm into the lymph or into the blood.

Even after the antigen has been destroyed, the plasma cells still continue to form antibodies. The reason for this is that once an antigen has been inside a cell, the cell's enzyme system has become permanently altered to produce the specific antibodies. Therefore, a single exposure to a toxic antigen sometimes makes the person immune to this toxin for long periods of time thereafter. In fact, when the plasma cells or other cells producing antibodies undergo mitosis to form new cells, each of the daughter cells also is still capable of producing antibodies, and this heritage continues for at least several generations of cells. However, the rate of antibody production progressively decreases over a period of months to a few years.

Specificity of immune bodies. Ordinarily, antibodies formed against a particular antigen will react only with that antigen. For instance, antibodies that protect against tetanus toxin are generally of value only for protection against this single toxin and not against any other type. Occasionally, however, toxins or bacterial proteins are sufficiently alike that antibodies formed against one will react against a second similar toxin or protein. In this instance, it is said that *cross-immunity* has developed between the two different agents.

Antigen-Antibody Reactions

Antibodies can protect against disease or toxins in a number of different ways. Sometimes the antibodies react directly with the toxins and destroy their toxicity. At other times they cause

the noxious agent to precipitate, agglutinate, rupture, or be phagocytized by white blood cells. The mechanisms of these reactions are the following:

Neutralization. The method by which most toxins are prevented from damaging the body's tissues is *neutralization.* Figure 87 illustrates a toxin with five separate *reactive areas* indicated by the large black dots. Each of these could cause toxic damage to the body's tissues. An antibody has attached itself to the antigen at each one of these reactive areas, thus preventing the reactive areas from damaging the tissue cells.

Precipitation. A similar method for preventing damage by toxins or other antigens is to cause them to *precipitate.* In this case the antibodies, which usually have two reactive areas, form bridges between the antigen molecules, connecting large numbers of antigens together. When the aggregate becomes large enough, it forms a precipitate and is usually phagocytized and digested by the reticuloendothelial cells.

Agglutination of organisms. The antigens of bacteria are part of the bacterial cell itself, and the antibodies attach themselves to the antigens on the surfaces of the bacteria. Since antibodies normally have two reactive areas, they can form bridges between adjacent bacterial bodies, causing large aggregates of bacteria to clump together. This process is called *agglutination.* Agglutinated bacteria are unable to spread through the body tissues and eventually are phagocytized.

Lysis of organisms. Antibodies can also cause bacteria to become *lysed,* which means rupturing of the cell with spillage of its contents. Lysis begins with many antibodies attaching themselves to protein antigens of the bacterial membrane. Then a substance called *complement,* a fatty material in the extracellular fluid, attaches to the other end of each antibody. Attachment of the antibodies by themselves does not cause lysis of the cells, but once complement attaches, the fat of the complement dissolves the cell membrane and causes the bacterium to break wide open.

Opsonization and phagocytosis of bacteria. Occasionally, when antibodies attach to the surface of the bacterium, the attachment of complement to the other ends of the antibodies still will not cause rupture of the cell. Yet, the presence of complement causes the bacterium to become especially susceptible to phagocytosis by the reticuloendothelial cells. The probable reason for this is that the complement, being a fatty substance, is miscible with the lipid membrane of the phagocytes. This process is called *opsonization,* meaning increased susceptibility of the bacterial cells to phagocytosis.

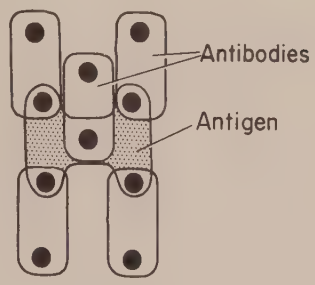


Figure 87. Neutralization of an antigen by antibodies.

Tissue Immunity Versus Humoral Immunity

Humoral immunity means immunity resulting from antibodies flowing freely in the body fluids. However, antibodies can also remain inside cells rather than enter the circulating body fluids. This results in a type of immunity called *tissue immunity.* The body retains tissue immunity many times as long as it retains humoral immunity. For instance, immunity to such infectious agents as smallpox virus may last for as long as ten years to a lifetime. One of the reasons for this seems to be the following: The cells that form the humoral immune bodies are the plasma cells and other reticuloendothelial cells that have life spans of only a few days. With each new generation of these cells, the rate of antibody formation decreases. Therefore, the quantity of humoral antibodies also decreases. On the other hand, many tissue cells, such as neuronal cells, muscle cells, and others, have extremely long life spans, and since these cells, too, can probably produce antibodies that remain in their cytoplasm, they can maintain immunity against noxious agents sometimes for the entire life of the individual. Thus, humoral immunity often cannot be demonstrated in a person, and yet he may be immune to tissue damage because of antibodies within the cells.

Active Immunity Versus Passive Immunity

The development of antibodies in response to a specific antigen is called *active immunity.* However, a person can also be immunized against an antigen by injection of plasma containing antibodies obtained from another human being or even from an animal that has been actively immunized against the antigen. These injected antibodies usually circulate in the recipient's blood for three to five weeks and then are destroyed by the reticuloendothelial system. This type of immunity is called *passive immunity.*

Passive immunity has an advantage over active immunity in acute illness, for it can be used clinically to increase the immunity of a person within minutes, whereas the person often will not develop sufficient active immunity to combat a disease for a week or more, in which case the disease frequently will have killed the person before the immunity can be of value. For example, in the case of a snake bite, the person is likely to be dead in only a few hours unless he is treated with immune sera. Immune sera used for instituting passive immunity are usually developed in animals, most frequently in horses.

Vaccination

A method for disease prevention familiar to everyone is *vaccination.* Vaccination causes development of antibodies against specific disease organisms or dangerous toxins without exposing the person to the disease. For instance, injection of dead bacteria that still contain antigenic proteins can cause immunity which later will assure attack of the live organisms equally as well as the dead organisms. The bacteria can be killed by heating them, by destroying them with formalin, destroying them with phenol, grinding them to bits, or even breaking them to pieces with high intensity sound waves.

Vaccines can be prepared from toxins by exposing them to chemicals that change their structure very slightly, enough to remove their toxic properties but not enough to destroy their antigenic properties. This can frequently be done by exposing the toxins to formalin, phenol, or some other similar chemical agent. The resulting substance is called a *toxoid.*

A third type of vaccine employs a living agent. For instance, smallpox virus is very similar to cowpox virus. Yet, cowpox virus causes only a single pustule to develop rather than multiple pustules all over the entire body. A person innoculated with cowpox virus develops cross-immunity to smallpox after the cowpox pustule has disappeared. Likewise, non-virulent strains of polio virus, yellow fever virus, and many other viruses have been used to cause immunity without causing the actual disease, or at least without causing a dangerous form of the actual disease.

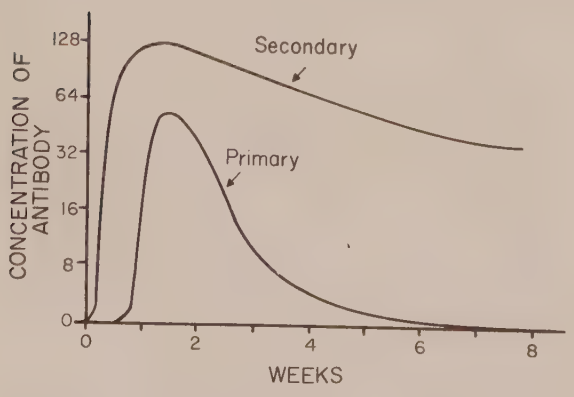


Figure 88. Antibody responses to successive injections of vaccine.

When a person has never been exposed to an antigen previously, either by injection of a vaccine or by exposure to the disease itself, the immune process usually takes one to two weeks after vaccination to build up a significant number of antibodies in the circulating body fluids, as shown by the lower curve of Figure 88. Then, over a period of another four to six weeks, the antibodies disappear. This is called the *"primary" immune response.* When the person is exposed to the same antigen a second time, whether it be a few weeks later or even many years later, the immune response develops in one to two days and is far more lasting; this is called the *secondary immune response,* which is illustrated by the upper curve of Figure 88.

ALLERGY

The Immune Reaction as the Basis of Allergy

When an antigen combines with an antibody while in direct contact with cells, the reaction usually damages the cells at the same time, causing swelling or even destruction of cells, followed by release of several different substances, including *histamine, heparin, leukotaxine, proteolytic enzymes,* and so forth, as illustrated in Figure 89. These substances then diffuse into the surrounding fluids and throughout the body to cause secondary reactions. The substance that causes the most serious effects is histamine, as is discussed later.

Weak immunity as a cause of allergy. As pointed out previously in relation to the immune reaction, most antibodies are "bivalent," which means that they have two reactive sites for combining with antigens. However, when antibodies are formed, one of these valencies is formed first and the other later. Sometimes the plasma cells empty univalent antibodies into the body fluids before they have formed the second reactive site.

This is particularly true when a person has been exposed to very weak antigens that are poor initiators of the immune process. Univalent antibodies have a special propensity for attaching themselves to tissue cells throughout the body. Therefore, there is far greater chance of immune reactions (and allergic reactions) taking place in contact with tissue cells when univalent antibodies are present than when bivalent antibodies are present.

ALLERGENS. The human body is continually exposed to many substances that cause very weak immunity but not strong immunity. These include such substances as partial proteins, conjugated proteins, or drugs that can combine weakly with proteins in the body. Among the most common of the poorer antigens are *pollens, industrial chemicals, animal dander, feather protein,* certain *food extracts,* and *drugs.* Since these have a tendency to cause allergy, especially in those persons who are normally predisposed to allergy, they are called *allergens.*

PREDISPOSITION OF CERTAIN PERSONS TO ALLERGY—THE ALLERGIC PERSON. The allergic person has a hereditary tendency to produce a higher proportion of univalent antibodies than is true in the normal person. This obviously predisposes such a person to allergy.

Different Types of Allergic Reactions

Primary allergies. Many different types of allergic reactions can occur, some of the more important types of which are the following:

ANAPHYLAXIS. Anaphylaxis results from injection of an antigen intravenously into a person who has previously been strongly immunized against the antigen. The direct reaction of the antigen and antibody in the blood stream and in tissues closely adjacent to the blood causes intense cellular damage and release of large quantities of histamine into the circulation. The histamine causes the capillaries to become very permeable to fluid so that plasma leaks into the surrounding tissues, thereby reducing the blood volume, and it causes the blood vessels to become markedly dilated, which causes pooling of large quantities of blood in the peripheral vasculature. The net result is a drastic fall in cardiac output, a condition called *anaphylactic shock,* that can cause death in a few minutes. Also, other tissues closely allied to the circulatory system become seriously damaged. For instance, bronchiolar spasm in the lungs frequently makes it very difficult for the person to breathe.

URTICARIA. Urticaria means the development of "hives" on the skin. It results from much the same type of reaction as that which causes anaphylaxis except that it takes place in many small localized areas throughout the body rather than everywhere, presumably at points where large quantities of the univalent antibodies are attached to the tissues. At these places the skin first becomes hyperemic (red because of excessive blood flow) because of arteriolar dilation. This is followed by increased leakage of fluid out of the capillaries into the tissues, resulting in large welts all over the surface of the body. These effects result mainly from the release of histamine by the damaged cells; antihistaminic drugs are extremely effective in preventing urticaria.

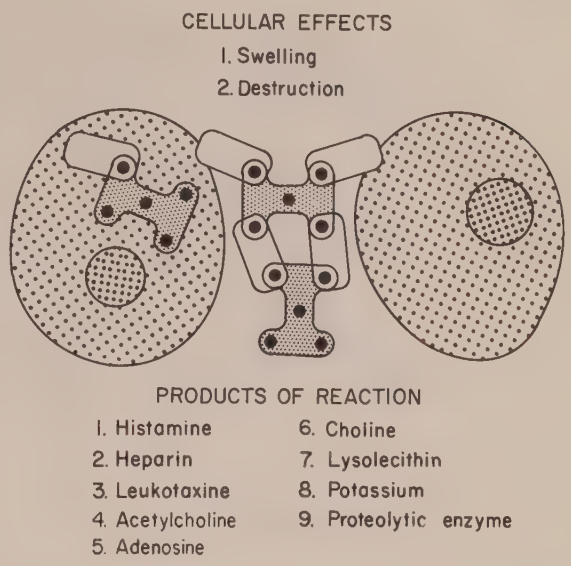


Figure 89. Cellular damage and substances released from the cell following an antigen-antibody reaction in contact with the cells.

HAY FEVER. Hay fever results from antigen-antibody reactions in the nose that probably cause both direct local tissue damage and histamine induced secondary damage. The histamine causes arteriolar dilatation in the local blood vessels, with resultant high capillary pressure and leakage of fluid into the tissues. The tissues become so edematous that the person has difficulty breathing through the nose.

ASTHMA. Asthma results from bronchiolar constriction, usually following respiratory exposure to antigens to which the person is sensitive. Antigens such as ragweed pollen, horse's dander, or even dust in the air can be breathed into the lungs where localized antigen-antibody reactions occur. This in turn causes intense bronchiolar constriction, often making it almost impossible for the person to breathe and almost suffocating him.

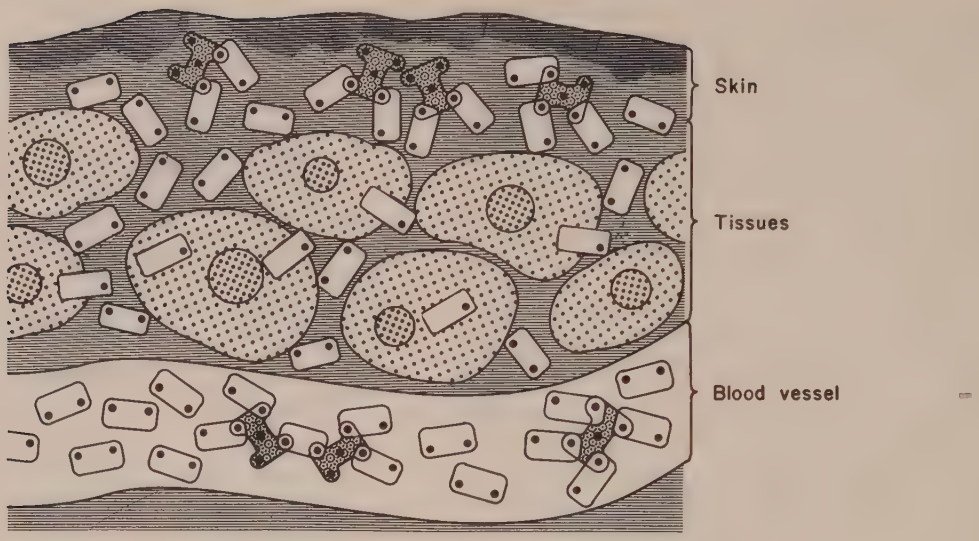


Figure 90. Effect of hyperimmunity in protecting against allergic reactions, showing attachment of antigens to bivalent antibodies before they can reach the univalent antibodies attached to the cells.

Desensitization (Hyperimmunization) as a Physiological Means of Treating Allergy

Most types of allergy, with the exception of anaphylaxis, are caused by weak immunity to an allergen, in which case the person has many univalent antibodies spread throughout his tissues, a state that predisposes to serious allergy. The person can frequently be made non-allergic by the process called *desensitization.* This is accomplished by injecting minute quantities of the allergen either intracutaneously or subcutaneously, a small amount every few days, until the person builds up increased immunity to the allergen. Once this has occurred, enough bivalent antibodies float around in the body fluids to destroy the allergens before they can ever reach the univalent antibodies attached to the tissue cells (see Figure 90). Thus, an allergic reaction is prevented.

REFERENCES

Ciba Foundation, Symposium Volume: *Cellular Aspects of Immunity.* Boston, Little, Brown and Company, 1960.

Henry Ford Hospital Symposium: *Mechanisms of Hypersensitivity.* Boston, Little, Brown and Company, 1959.

Kabat, E. A., and Mayer, M. M.: *Experimental Immunochemistry.* Springfield, Ill., Charles C Thomas, 1958.

Lawrence, H. S.: *Cellular and Humoral Aspects of the Hypersensitive States.* New York, Paul B. Hoeber, Inc., 1958.

Mongar, J. L., and Schild, H. O.: Cellular mechanisms in anaphylaxis. *Physiol. Rev.* 42:226, 1962.

Najjar, V. A.: Some aspects of antibody-antigen reactions and theoretical considerations of the immunologic response. *Physiol. Rev.* 43:243, 1963.

Raffel, S.: *Immunity; Hypersensitivity; Serology.* New York, Appleton-Century-Crofts, 1953.

Smith, D. T., and Conant, N. F.: *Zinsser's Bacteriology.* 11th Ed. New York, Appleton-Century-Crofts, 1957.

Talmage, D. W., and Cann, J. R.: *The Chemistry of Immunity in Health and Disease.* Springfield, Ill., Charles C Thomas, 1960.

BLOOD COAGULATION, TRANSFUSION, AND TRANSPLANTATION OF ORGANS

11

BLOOD COAGULATION

Blood coagulation prevents blood loss when vessels are ruptured. After a vessel begins to bleed, physical and chemical processes are set into operation, leading in approximately two to six minutes to the formation of a clot. The clot plugs the rent in the vessels and prevents further loss of blood.

The Mechanism of Blood Coagulation

The general scheme of blood coagulation is shown by the following physical and chemical reactions:

1. The traumatized tissues at the edge of the broken vessel release *tissue thromboplastin.* Platelets also adhere to the broken edges of the vessel as shown in Figure 91 and then disintegrate to cause formation of *plasma thromboplastin.*
2. *Thromboplastin + calcium + prothrombin + accelerator globulin + proconvertin + Stuart factor* all enter into a series of reactions that form *thrombin.*

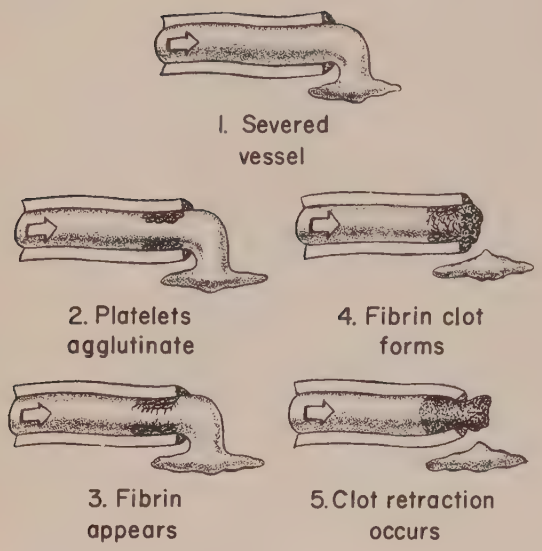


Figure 91. Stages of clot development and clot retraction following injury to a blood vessel. (Redrawn from Seegers and Sharp: *Hemostatic Agents.* Charles C Thomas.)

3. *Thrombin + fibrinogen → fibrin.*
4. *Fibrin threads* entrap red blood cells to form the *clot.*

Thromboplastin and its formation. There are several different types of thromboplastin composed principally of lipid materials. These lipid materials are present in all tissues, and damage to the tissues causes their release into the surrounding fluids. Thus, trauma that causes a blood vessel to rupture obviously damages the tissues along the edges of the rupture, releasing *tissue thromboplastin* and thereby immediately starting the clotting process.

Thromboplastin is also formed from platelets but not in so simple a manner, because thromboplastin is not present in the platelets in the form of thromboplastin itself. Disintegration of the platelets releases certain lipid materials, especially the phospholipid *cephalins,* which initiate a series of reactions with plasma proteins to form a type of thromboplastin called *plasma thromboplastin.* The different substances that enter into this reaction are: (1) *calcium,* (2) *antihemophilic factor,* (3) *plasma thromboplastin component,* (4) *plasma thromboplastin antecedent,* (5) *Hageman factor,* (6) *Stuart factor,* and (7) *accelerator globulin.* Thus, it is readily evident that this is a very complicated reaction requiring calcium ions and a number of different protein factors from the blood plus a lipid material from the platelets. The substance that is finally formed, plasma thromboplastin, can then act in the same way as tissue thromboplastin to promote blood clotting.

Platelet disintegration. *Platelets* are 1- to 2-micron fragments of large *megakaryocytes,* a type of white blood cell formed in the bone marrow. The number of platelets in each cubic millimeter of blood is normally about 400,000. Platelets have a thin membrane that is negatively charged on the outside. The endothelial walls of the vessels throughout the circulation are also normally negatively charged so that platelets are repelled from the walls of these vessels. However, when a vessel ruptures, the damaged endothelial cells lose their negative charge, and platelets then begin to stick at the point of rupture. Likewise, platelets can stick to foreign bodies in the circulation or even to a forming clot. In any of these instances, the process of adhering to some other object causes the platelets to disintegrate and dump their substance into the surrounding fluids. Among the expelled substances are lipids that can initiate the formation of plasma thromboplastin, as discussed in the preceding paragraph.

Formation of thrombin. Thromboplastin, having been released into the blood from the tissues, or having been formed as a result of platelet disintegration, acts as an enzyme to initiate the second of the above reactions in the clotting process. For this reaction to take place, *ionic calcium,* and several proteins, *accelerator globulin, Stuart factor, proconvertin, prothrombin,* and perhaps other less important factors all enter together into a chemical reaction to produce the substance *thrombin.* This reaction takes place in only a few seconds once an adequate amount of thromboplastin is available.

Formation of fibrin. Once thrombin has been formed, it causes *fibrinogen* to change into *fibrin.* Fibrinogen is one of the plasma proteins usually present in the plasma in a concentration of about 0.3 per cent. Thrombin acts as an enzyme to activate the fibrinogen molecule, forming what is called *fibrin monomer,* which means a single molecule of activated fibrin. These molecules automatically polymerize with each other into long *fibrin threads* forming a *recticulum* that entraps red blood cells, white blood cells, and platelets to form a *clot.*

Clot retraction. Once the clot has formed, the fibrin threads gradually contract, expressing plasma out of the clot but keeping the red blood cells, platelets, and other solid elements. The plasma that is extruded from the clot has no fibrinogen in it because this has been converted into fibrin. Therefore, it is called by another name, *serum.*

Fibrin threads attach themselves to the

damaged surfaces of the blood vessels. Therefore, as the blood clot retracts, it pulls the edges of the opening together, thus contributing to ultimate hemostasis.

For clot retraction to occur to a maximum extent, an adequate number of platelets must be present in the blood. Electron micrographs of blood clots show that platelets form nidi for development of more and more new fibrin threads, which in turn contribute to the contraction of the clot. It is presumed that the platelets that become enmeshed in the clot slowly disintegrate, thereby causing the formation of more and more plasma thromboplastin, which in turn leads to more and more formation of fibrin.

Organization of the blood clot. After a clot is formed, its immediate function of stopping blood loss will have been achieved. To effect permanent repair of the blood vessel, however, the clot becomes *organized,* which means ingrowth of fibrous cells to form connective tissue. This process continues for several days, and endothelial cells simultaneously grow over the vascular surface of the clot to form a new smooth lining on the inside of the vessel. Thus, after only a few days, the break in the vessel not only will have been repaired but also will have been provided with a new inner lining.

Vascular spasm as an aid to hemostasis. The term *hemostasis* means the arrest of an escape of blood. The blood clotting mechanism is the most important of all of the hemostatic mechanisms, but another one that is of immediate importance after a vessel is ruptured is spastic contraction of the damaged vessel itself. Trauma to a vessel initiates impulses that travel along the surfaces of the smooth muscle fibers in the vascular wall. These impulses cause contraction of the muscle, with resultant spastic closure of the vessel. This prevents serious bleeding even before the clot can be formed. Indeed, ruptured arteries as large as three-eighths inch in diameter, and veins even larger, sometimes contract so spastically that the person will not bleed to death. Without such contraction enough blood could be lost in less than 20 seconds to cause death.

Prevention of Intravascular Clotting Under Normal Conditions

Since thromboplastin and thrombin are both enzymes, it would theoretically be possible for them to cause continual coagulation in the circulation once even a small amount of each of them has been formed. Fortunately, this does not occur for many different reasons, some of which are the following:

Endothelial repulsion of platelets. The first method for preventing intravascular clotting is repulsion of platelets away from the endothelial linings of the blood vessels caused by electronegative charges on both the endothelial wall and platelets as described above. It is believed that special proteins adhere to the vessel lining with their molecules so oriented that strong negative electrical charges protrude outward into the blood vessel. Theoretically, as long as this protein lining remains intact, the platelets cannot adhere to the surface of the vessel wall and, therefore, cannot initiate a clot.

Antithromboplastin. A second method for preventing clotting is destruction of thromboplastin. Thromboplastin loses its activity in the blood within ten minutes to an hour. Therefore, it is said that plasma has *antithromboplastin activity,* though the exact substance that is responsible for this is yet unknown. Obviously, then, thromboplastin released into the blood cannot continue to promote more and more clotting indefinitely.

Antithrombin effects. Thrombin is also removed from the blood in several different ways. First, a large portion of the thrombin is directly adsorbed to the fibrin threads of the clot as they are formed. Second, a special protein in the plasma called *antithrombin* acts enzymatically on thrombin to inactivate it, and, third, a substance called *heparin,* which is discussed below, probably also helps to destroy the thrombin.

Heparin. Another means to prevent intravascular clotting is continual secretion into the blood of the anticoagulant *heparin.* This is a conjugated polysaccharide that is secreted into the circulatory system continually by *mast cells* located in the pericapillary connective tissue throughout the body. These cells appear to be typical basophils as shown in Figure 92, one of the types of white blood cells described in Chapter 9. It is believed that *basophils* circulating in the blood can also secrete heparin, though this has not been proved.

Heparin blocks the sequence of reactions that converts prothrombin into thrombin. One of the ways it does this is to form a complex consisting of heparin, accelerator globulin, and antithrombin, in this way reducing the amount of accelerator globulin available to activate prothrombin. It has also been claimed that heparin breaks up some of the plasma lipoproteins and in the process releases a powerful phospholipid anticoagulant, *sphingosine,* which inhibits the conversion of prothrombin to thrombin. Regardless of the precise means by which heparin acts to prevent intravascular clotting, its concentration in the blood is sufficient under normal conditions to prevent the

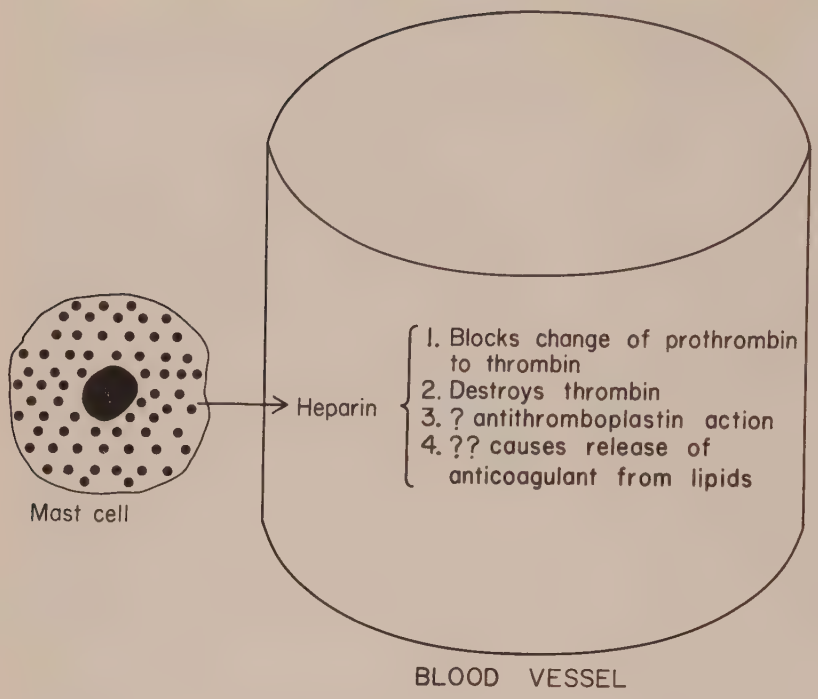


Figure 92. Normal prevention of intravascular clotting by heparin released from mast cells and basophils.

formation of significant amounts of thrombin in the circulating blood and to remove that small portion of thrombin that does happen to be formed.

Extravascular anticoagulants. Some anticoagulants are used artificially to prevent blood coagulation outside the body. The most important of these are chemicals that remove calcium ions from the blood. *Oxalate salts,* for example, when added to blood cause precipitation of calcium oxalate, removing the calcium ions and blocking the clotting process.

Citrate salts also remove calcium ions from the blood by forming calcium citrate, a soluble substance but almost totally non-ionized. Blood that has been rendered incoagulable by citrate can still be reinjected into a person provided that it is given slowly, because almost as rapidly as it enters the circulation the citrate is destroyed by the liver. The blood again becomes capable of coagulating and is relatively normal in all other respects. Therefore, in giving transfusions, blood is usually drawn from a donor and mixed immediately with a mixture of sodium, potassium, and ammonium citrate to prevent coagulation. Upon readministration the citrate is destroyed.

Bleeding Diseases

Lack of any one of the many different blood clotting factors can prevent adequate blood coagulation, so that once blood vessels rupture blood continues to ooze for many hours. A large number of blood vessels rupture internally each day, and were it not for the blood clotting process, a person would actually bleed to death internally. Fortunately, in most bleeding diseases, clotting does occur to some extent but much more slowly than normally.

Hemophilia. There are several different types of hemophilia, all of which are hereditary but each of which has slightly unique characteristics. These diseases are caused by lack of *antihemophilic factor, plasma thromboplastin component,* or *plasma thromboplastin antecedent,* three substances that enter into the formation of plasma thromboplastin. However, these three factors are not necessary for the formation of tissue thromboplastin. Therefore, if the tissues of a hemophilic person are torn severely enough to form adequate amounts of tissue thromboplastin, clotting occurs almost normally, and very little blood is lost. On the other hand, simple rupture of a blood vessel does not form much tissue thromboplastin, in which case bleeding often lasts for hours. The gradual oozing of blood frequently allows the hemophiliac to bleed to death.

Thrombocytopenia. Platelets are also called *thrombocytes,* and a deficiency of thrombocytes is called *thrombocytopenia.* An occasional person develops a deficiency of platelets down to less than one-tenth the normal level. In these persons, plasma thromboplastin cannot be formed, so that

internal bleeding occurs very frequently. However, an additional effect besides those that take place in hemophilia also occurs. This is oozing of blood from the minute capillaries. It is believed that very small breaks in capillaries are blocked not by clots but instead by direct adherence of platelets to the rents in the vessels. In platelet deficiency, minute rents in the very small vessels cause literally thousands of small bleeding areas throughout the body only a few millimeters in diameter. Thus, the person becomes studded with many minute hemorrhagic spots beneath his skin. Bleeding of this same type also occurs internally in all the organs.

Decreased prothrombin or proconvertin—vitamin K deficiency. Two other clotting factors frequently absent or deficient in the circulation, prothrombin and proconvertin, are proteins formed by the liver, and both require vitamin K as one of the intermediary substances in their formation. In vitamin K deficiency, the quantities of these substances in the blood become very deficient. Also, diseases of the liver, such as hepatitis, cirrhosis, or acute atrophy of the liver, can depress the rate of prothrombin or proconvertin formation.

When either prothrombin or proconvertin is deficient in the blood, the formation of thrombin is greatly depressed so that a bleeding tendency develops.

Blood Coagulation Tests

Bleeding time. Several tests are used to determine the integrity of the blood coagulation process. The most common of these is a simple bleeding test in which a finger or a lobe of an ear is pricked and the time required for the bleeding to stop is measured. Following an ordinary prick of the finger, bleeding stops in 2 to 3 minutes, and bleeding of the lobe of the ear normally stops in 4 to 5 minutes.

Clotting time. Another test is to determine whether or not blood removed from the person will coagulate as rapidly as normally. To do this, blood is collected in an absolutely clean test tube and observed until it clots. Ordinarily, the clotting time is about 6 minutes if the blood has not been agitated too much during the test.

Platelet count. If blood fails to coagulate in the test tube in a reasonable period of time, one searches for the factors that might be missing in the coagulation process. One of the first tests is to count the number of platelets. As long as the number of platelets is above 150,000 per cubic millimeter, the person will have no clotting difficulty as a result of platelet deficiency. However, in persons with thrombocytopenia, the platelet count frequently falls to as low as 40,000 to 50,000, in which case excessive bleeding does occur.

Prothrombin time. A test to determine the quantity of prothrombin in the blood is performed by first rendering the blood incoagulable with oxalate which precipitates the ionic calcium. Then large amounts of calcium, thromboplastin, accelerator globulin, and proconvertin are all mixed with the sample of blood suddenly. The only additional factor needed for the formation of thrombin that is not added to the blood is prothrombin. Therefore, the length of time required for the blood to clot depends on the amount of prothrombin already in the blood. The longer the time required, the less the amount of prothrombin available. The normal prothrombin time is approximately 12 seconds—that is, this is the time required for the clot to appear after mixing has occurred. If the prothrombin time is as long as 20 to 30 seconds, the person can be expected to be a bleeder.

The quantities of proconvertin, accelerator globulin, or any of the other factors that enter into blood coagulation can be estimated in the same way.

TRANSFUSION

Often a person loses so much blood that to save his life he must be given a transfusion of new blood immediately. At other times transfusions are given to treat anemia or some other blood deficiency. Unfortunately, the bloods of different people are not all exactly alike, and failure to inject the appropriate type of blood is likely to cause death of the recipient. For this reason it is very important to understand the physiology of blood grouping and blood matching prior to giving transfusions.

Blood Matching

The reason one person's blood may not be suitable for another person is that the recipient may be immune to some of the proteins in the blood cells of the donor. Immune bodies in the recipient's plasma can cause *agglutination* and *hemolysis* of the injected cells, thus plugging some of the vessels and releasing large quantities of hemoglobin out of the cells into the circulation.

Cross-matching. One means by which the suitability of blood for transfusion can be determined is to *cross-match* the bloods of a donor and

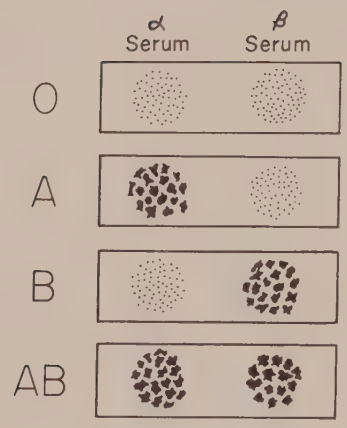


Figure 93. Agglutination of some cells and lack of agglutination of other cells in the process of blood typing.

recipient. To do this, the two bloods are both centrifuged, and the plasma and red cells separated from each other. Then the red blood cells are resuspended in a saline solution. A small quantity of cells from the donor is mixed with a small quantity of plasma from the recipient, and similarly a small quantity of cells from the recipient is mixed with a small quantity of plasma from the donor. These are allowed to remain together for 3 to 15 minutes, and then are observed under a microscope. If none of the cells are agglutinated as shown by the clumps on the microscope slides in Figure 93, it can be assumed that the two bloods *match* each other, or, in other words, that no immune bodies are present in either of the two bloods that can damage the cells of the other blood. However, if on microscopic examination clumps of red cells are evident, then it is obvious that an immune reaction has occured and that the bloods are mismatched.

The blood groups—types A, B, and O. The different types of blood are classified on the basis of the *cell proteins* that normally cause immune reactions. The two proteins in the cells that most frequently cause reactions are the type A and type B proteins; these are called *agglutinogens*. All human bloods can be classified into four major groups, which are given in Table 1. These are groups O, A, B, and AB. The red cells of group O blood contain neither A nor B agglutinogen. Group A red blood cells contain type A agglutinogen, and group B red blood cells contain type B agglutinogen. Finally, group AB cells contain both A and B agglutinogens.

Table 1. *The Blood Groups*

Blood Type	*Agglutinogens*	*Agglutinins*
O	O	α, β
A	A	β
B	B	α
AB	A, B	none

In a person's *plasma* are immune bodies that agglutinate cells containing agglutinogens different from those in his own cells. These immune bodies are called *agglutinins*, and alpha (α) agglutinins agglutinate type A cells, while beta (β) agglutinins agglutinate type B cells. Thus, as shown in Table 1, a person with type O blood has both alpha and beta agglutinins in his plasma. These are capable of agglutinating cells containing either type A or B agglutinogens. Type A blood contains beta agglutinins but not alpha agglutinins, and type B blood contains alpha agglutinins but not beta agglutinins. Finally, type AB blood contains no agglutinins whatsoever because both of the agglutinogens are present in the cells.

BLOOD TYPING. To determine the type of blood, a procedure called *blood typing* is performed. The red cells from the subject are mixed with two samples of plasma, one containing large amounts of alpha agglutinins and the other large amounts of beta agglutinins. Figure 93 illustrates the reactions observed on a microscope slide with each type of blood when this is performed. In the case of type O cells, none of the cells are agglutinated by either the alpha or beta agglutinins because these contain neither A nor B agglutinogens. In the case of type A cells, alpha agglutinins agglutinate the cells but beta agglutinins do not because these cells contain A agglutinogens. In the case of type B cells, beta agglutinins agglutinate the cells but alpha agglutinins do not, and in the case of type AB cells, both alpha and beta agglutinins agglutinate the cells. In other words, by observing which agglutinins agglutinate the cells, one can determine the presence or absence of A or B agglutinogens in the cells and thereby determine the blood type of the subject.

The Rh factor. In addition to the four major blood groups, another protein, the *Rh factor*, present in the red cells of approximately 85 per cent of all persons, sometimes causes mismatching of blood. However, the Rh factor is different from the A and B agglutinogens in the following way: Those persons who do not have Rh factor (the Rh negative persons) do not normally have Rh antibodies. This is different from the A and B blood groups, for those persons who do not have either A or B agglutinogens do have respective agglutinins for agglutinating these substances. Yet, once blood cells containing the Rh factor have been injected into a recipient who does not have the

Rh factor, the recipient forms Rh antibodies. On second injection of the Rh cells the Rh antibodies cause agglutination and a typical transfusion reaction.

To recapitulate, the Rh factor cannot cause a transfusion reaction on first exposure of mismatched blood, but once an *Rh negative person* has been immunized against Rh positive blood, second exposure to the Rh positive blood causes typical agglutination.

ERYTHROBLASTOSIS FETALIS CAUSED BY THE Rh FACTOR. One of the most important consequences of the Rh factor is the abnormality called *erythroblastosis fetalis,* which ocassionally occurs in newborn infants. It occurs in about 1 out of every 50 children born to Rh negative mothers when the father is Rh positive. If a fetus inherits the Rh positive characteristic from the father, its red cells will contain the Rh factor. Many of the fetus' red blood cells continually disintegrate, and some of the Rh protein passes through the placental membrane into the mother. This causes the mother to develop Rh antibodies, which in turn pass back through the placental membrane into the baby, resulting in agglutination and destruction of the baby's red blood cells. To offset this rapid destruction, the baby's bone marrow, spleen, and liver produce red blood cells extremely rapidly, so rapidly, in fact, that many of the cells emptied into the circulation are early, nucleated red blood cells. This is the reason why the disease is called "erythroblastosis."

Many infants with erythroblastosis are born dead because of the inability of their anemic blood to provide necessary oxygenation of the tissues. Those that are born alive are likely to die within the first few weeks. The standard treatment is to transfuse the child with Rh negative blood while bleeding him at the same time. In this way his Rh positive blood is gradually removed and is replaced by Rh negative blood that cannot be destroyed by the Rh antibodies in the child's body fluids. Within three to five weeks after birth the Rh antibodies will have been destroyed, and the child's own Rh positive cells will then no longer be destroyed.

Other factors that lead to mismatching of blood, and their use in forensic medicine. Often present in the red cells are several other proteins that can cause transfusion reactions in the same manner as the Rh factor. Some of these are Hr, M, N, P, Lewis, and Kell factors. Ordinarily, however, these substances are so poorly antigenic that the transfusion reactions that result are very slight. The major importance of these factors is that they are often used in forensic medicine to determine familial relationships between persons. On the average, half of the factors found in a child's blood will also be found in the father's blood and the other half in the mother's blood. Thus, by a process of elimination, it is sometimes possible to disprove fatherhood in cases of disputed parentage.

Techniques of Blood Transfusion

For transfusing blood any method can be used that will transfer the blood from a donor to a recipient without causing coagulation or destruction of the blood in the process. The earliest method was simply direct connection between an artery of the donor, by means of needles and a rubber tube, and a vein of the recipient. Blood was allowed to flow from the donor to the recipient until the donor fainted or until the recipient showed signs either of improvement or of a transfusion reaction.

The most usual technique now used for transfusion is to collect blood from the donor in a container while mixing the blood continually with a small quantity of sodium, potassium, and ammonium citrate and a small quantity of glucose. The citrate compounds prevent blood coagulation, and the glucose provides nutrition for the red cells and keeps them structurally sound. This blood can be given immediately to the recipient, or it can be placed in a refrigerator and kept at 4° C. for as long as three weeks prior to administration. This is the method used by hospital blood banks for storing blood until some patient needs a transfusion.

Effect of administering mismatched blood. When mismatched blood is administered to a recipient, the first effect is usually agglutination of the red blood cells. The cellular clumps then flow into the small vessels, as illustrated in Figure 94, and fail to pass on through the capillaries. This blocks blood flow in the respective vessels temporarily, but after a few minutes to a few hours the wedging of the cells further into the smaller vessels and the destructive effect of antibodies on the cellular membranes cause the red cells to disintegrate. This allows the blood to flow once again, but large quantities of hemoglobin are dumped into the plasma. The final result, therefore, is destruction of many red blood cells, with the release of hemoglobin into the circulation.

EFFECT OF TRANSFUSION REACTIONS ON THE KIDNEYS. Actually, the clumping of the cells and the loss of the cells themselves are rarely of great importance in transfusion reactions. Instead, it is the free hemoglobin released into the plasma that usually causes the damage. The hemoglobin molecule is small enough to pass through capillary pores, and much of it passes through the glomeru-

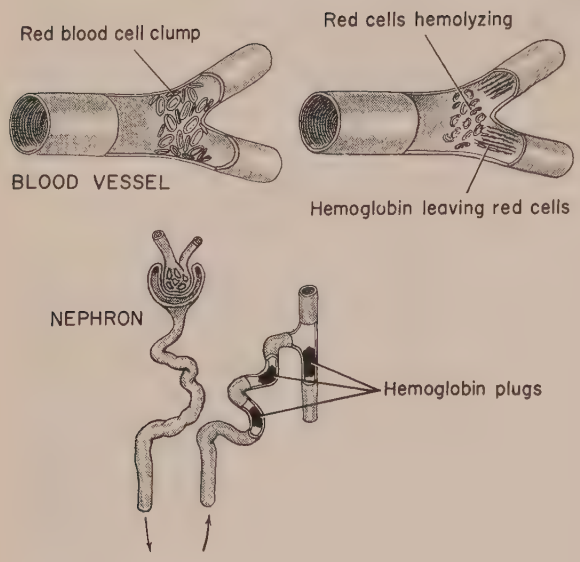


Figure 94. Mechanism of transfusion reactions caused by mismatched blood.

lar capillaries into the kidney tubules. As the tubules reabsorb water and electrolytes, the hemoglobin in the tubular fluid becomes progressively more and more concentrated. Finally, the concentration becomes so great that hemoglobin precipitates and plugs the tubules, as illustrated at the bottom of Figure 94. Obviously, a hemoglobin plug stops further function of the nephron. When this occurs in all the tubules of both kidneys, death due to renal shutdown is likely in 8 to 12 days.

Once a transfusion reaction has occurred, prevention of death is often possible by immediate institution of three physiologic procedures: First, administration of large quantities of water causes water diuresis, making it less likely for hemoglobin to precipitate in the tubules. Second, administration of alkaline substances causes the formation of an alkaline tubular fluid that can hold more hemoglobin in solution than can acidic fluid. Third, administration of diuretics also causes rapid flow of tubular fluids and helps to prevent hemoglobin plugs.

Usually, if death does not occur within two weeks after a transfusion reaction, automatic disintegration of the hemoglobin plugs in the tubules will allow many of the tubules to open up again and the kidneys to become functional.

Blood Substitutes

Many different solutions can be injected into the circulation to increase the blood or plasma volume. Solutions used for this purpose are called *blood substitutes.* Often a blood substitute can be used more advantageously than can blood itself. For instance, on the battlefield it is very difficult to maintain adequate quantities of different types of blood, but blood substitutes can be given regardless of blood type, and, therefore, permit rapid and easy use under stressful conditions. Several blood substitutes are the following:

Plasma. Plasma is separated from whole blood by centrifuging the red blood cells to the bottom of a container and decanting the plasma. Plasma can be used in place of whole blood to restore normal blood volume in cases of hemorrhage. Because of its protein content, it creates colloid osmotic pressure at the capillary membranes, and this keeps the fluid of the plasma in the circulation. Therefore, except for a resulting anemia, the circulatory system of a person who has hemorrhaged will be essentially normal again. Ordinarily, administration of plasma is completely adequate for treating hemorrhagic shock. However, when extreme quantities of blood have been lost, it is necessary to give at least some whole blood to restore the blood's oxygen-carrying capacity.

Saline. Saline, prepared in a 0.9 per cent solution by adding common salt to water, is isotonic with the body fluids, and will not cause osmotic changes in the red blood cells or other body cells. On intravenous injection this solution diffuses throughout the entire extracellular fluid compartment, approximately nine-tenths going into the interstitial spaces and approximately one-tenth remaining in the blood. Therefore, saline can be used in treating blood loss, though it causes some edema at the same time.

Other solutions. A blood substitute that has become very valuable on the battlefield is a solution of *dextran.* This substance is a large polysaccharide compound that has a molecular size greater than that of plasma proteins. These molecules, like those of plasma proteins, do not pass through the capillary pores and, therefore, also like plasma proteins, cause colloid osmotic pressure. As a result, dextran solutions have almost exactly the same properties as plasma. They remain in the circulation, increasing the blood volume back to normal while not causing any edema of the interstitial spaces. Further advantages of dextran solutions are that they cause no allergic reactions, they do not require citrate anticoagulant as does plasma, and they can be prepared in large quantities by commercial processes without necessitating removal of blood from donors.

SOLUTIONS FOR INTRAVENOUS NOURISHMENT. Nourishment is often given by intravenous methods. Solutions of glucose, amino acids, and even emulsified fats can be given intravenously

for this purpose. Occasionally plasma and whole blood transfusions are given simply because of their nutrient value, and not because of a need for enhanced blood volume.

SOLUTIONS FOR REESTABLISHING ELECTROLYTE EQUILIBRIUM. Occasionally, also, solutions are administered to reestablish electrolyte equilibrium. For instance, a person whose extracellular fluids contain too few potassium ions can be treated by administering a potassium solution. Or a person whose fluids are too acidic can be treated by administering sodium lactate, sodium gluconate, or sodium bicarbonate. The lactate and gluconate radicals are removed by metabolic processes in the liver, and the bicarbonate radical becomes carbon dioxide and is blown off through the lungs. As a result, the sodium ions alkalinize the acidic fluids.

TRANSPLANTATION OF TISSUES AND ORGANS

Before discussing transplantation of tissues or organs, we must first note that all cells of the body have antigenic proteins the same as those found in the blood cells. Thus, all cells of a person with type A blood have type A agglutinogens. Likewise, an Rh positive person has Rh positive antigen not only in his blood cells but in all his other cells. Among the many different types of antigenic proteins present in the cells of different persons are the A factor, the B factor, the O factor, three different Rh factors, three different Hr factors, and the M, N, S, P, Kell, Lewis, Duffy, Kidd, Diego, and Lutheran factors. Also, besides these well characterized factors still many others exist.

The antigenic factors of a transplanted tissue, if the tissue comes from another person, almost always initiate immune responses in the host, and within a few weeks enough antibodies are built up to destroy the transplanted tissue.

Autologous, homologous, and heterologous transplants. A transplant from one part of a person's body to some other part of his own body is called an *autologous* transplant. Such transplants can be skin, bone, cartilage, and so forth. Such grafts survive indefinitely, because the body will not build up antibodies against its own tissues.

A *homologous* transplant is one from one member of a species of animal to another member of the same species, such as a transplant from one human being to another. In this case, antibodies usually develop within 3 to 10 weeks in sufficient quantity to destroy the transplant. The one instance in which homologous transplants are successful is from one identical twin to another, for identical twins have precisely the same antigenic substances in their cells. Obviously, therefore, a transplant from one such twin to the second is the same as an autologous transplant.

A *heterologous* transplant is one from a member of one species of animal to a member of another species. In this case, the antigenicity of the transplanted tissue is so great that the tissue or organ is destroyed in two to three weeks.

Transplantation of non-cellular tissues. In a few instances, non-cellular tissues are transplanted from one person to another not for the purpose of providing new cells but for providing a framework into which the recipient's own tissues can grow. For instance, a bone graft from a horse to a human being frequently provides an appropriate latticework into which the human being can grow new bone. All the cells of the transplanted bone die so that only the physical structure of the bone is important. Another transplant that is usually successful is the cornea of the eye from one person to another. The clear part of the cornea has no living cells except for thin linings on the two surfaces, and new linings grow from the host's tissues at the edges of the cornea.

Function of the thymus in relation to transplants. Until very recently, physiologists knew no specific function for the *thymus gland.* Furthermore, removal of the thymus gland from an adult animal causes essentially no changes in growth, development, or function of an animal. However, if a thymus gland is removed during the first few months of fetal life, the animal fails to develop his mechanism for forming antibodies. Therefore, it is presumed that the thymus in the embryonic animal forms cells that spread by way of the blood to all parts of the body and eventually lodge in the lymphoid tissue to become the precursors of the plasma cells, the cells that in later life form almost all of the antibodies.

Furthermore, the cells of the thymus gland adapt themselves during fetal life so that they will not form antibodies against the animal's own tissues. Furthermore, if the fetus is exposed to certain foreign proteins continually during the first few months of fetal life, the immune system will never form antibodies against these same proteins throughout the life of the individual.

These recent studies on the thymus are mentioned here because they provide one of the most likely possibilities for successful transplantation of tissues and organs from one human being to another. Suppose, for instance, that a fetus is injected with a mixture of foreign proteins from many other human beings. The person originating from this fetus would then be unable to form antibodies against antigenic substances from other persons. Therefore, it should be possible to transplant tissues or organs into him at will.

REFERENCES

Biggs, R., and MacFarlane, R. G.: *Human Blood Coagulation.* 2nd Ed. Springfield, Ill., Charles C Thomas, 1957.

Brinkhouse, K. M.: Blood clotting: the plasma procoagulants. *Ann. Rev. Physiol.* 21:271, 1959.

Kabat, E. A.: *Blood Group Substances.* New York, Academic Press, 1956.

Kerr, D. J. A.: *Forensic Medicine.* 6th Ed. New York, The Macmillan Company, 1958.

Lawrence, H. S.: Homograft sensitivity. *Physiol. Rev.* 39:811, 1959.

Merrill, J. P.: Transplantation of normal tissues. *Physiol. Rev.* 39:860, 1959.

Race, R. R., and Sanger, R.: *Blood Groups in Man.* 3rd Ed. Springfield, Ill., Charles C Thomas, 1958.

Wiener, A. S.: *An Rh-Hr Syllabus.* 2nd Ed. New York, Grune & Stratton, Inc., 1958.

Section Four

THE CARDIOVASCULAR SYSTEM



THE PUMPING ACTION OF THE HEART, AND ITS REGULATION

12

The Heart as a Pump

The function of the heart is to keep blood flowing through the circulatory system. The heart is actually two separate pumps, as shown in Figure 95, one that pumps blood from the systemic circulation into and through the lungs, while the other pumps blood from the lungs through the remainder of the body and back again to the heart. Thus, the blood flows around and around a continuous circuit.

Figure 96 shows the functional details of the heart as a pump. Blood entering the right atrium from the large veins is forced by atrial contraction through the tricuspid valve into the right ventricle. The right ventricle pumps the blood through the pulmonary valve into the pulmonary artery, thence through the lungs, and finally through the pulmonary veins into the left atrium. Left atrial contraction then forces the blood into the left ventricle, from whence it is pumped through the aortic valve into the aorta and on through the systemic circulation.

The two atria are primer pumps that force blood into the respective ventricles. This propulsion of extra blood into the ventricles makes them more efficient as pumps than they would be if they had no special filling mechanisms. However, the ventricles are so powerful that they can still pump large quantities of blood even when the atria fail to function.

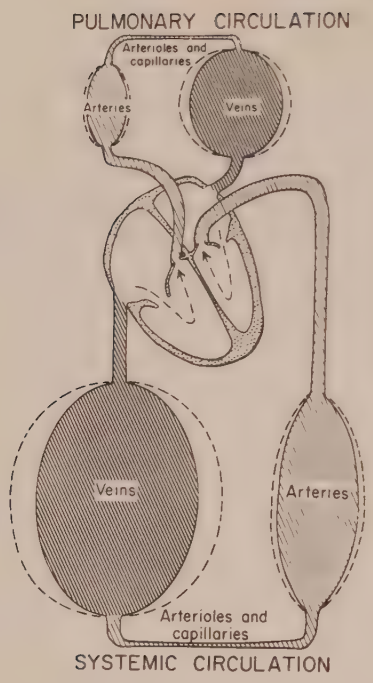


Figure 95. Schematic representation of the circulation, showing the two sides of the heart and the pulmonary and systemic circulations.

CARDIAC RHYTHMICITY AND ITS REGULATION

In order to pump blood the heart must alternately relax and contract, allowing blood to enter its chambers during the relaxation phase and forcing it out during the contraction phase. The alternate contraction and relaxation is provided by an inherent rhythmicity of the cardiac muscle itself, which is seen beautifully by simply observing an exposed heart. Recordings of the contractions of different portions of the turtle heart, for instance, are illustrated in Figure 97, which shows at the top a record from the sinus of the heart, in the middle a record from the atrium, and at the bottom a record from the ventricle. No nerves are attached to the different portions of the heart to make them contract, and no other signals arrive from outside the heart to cause the rhythmicity. In other words, the rhythmicity of the heart is vested in the heart muscle itself, and if any portion of the heart is removed from the body, it will continue to contract as long as it is provided with sufficient nutrition.

The sinus as the pacemaker of the heart.

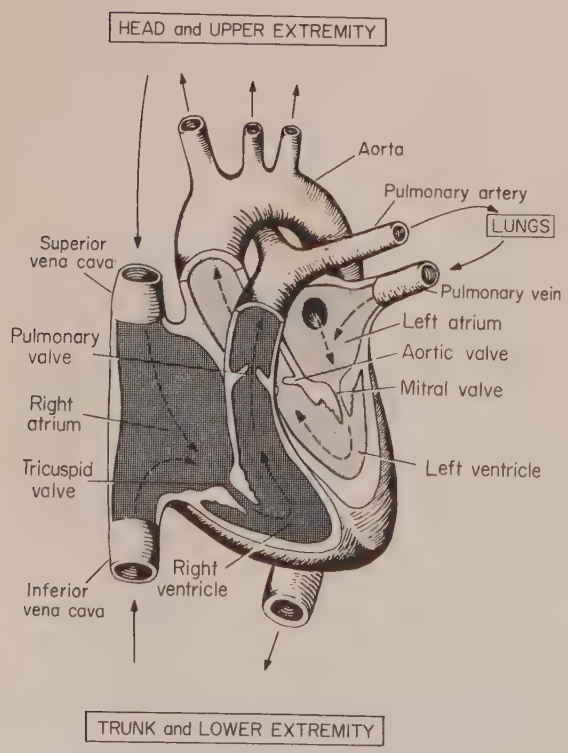


Figure 96. The functional parts of the heart.

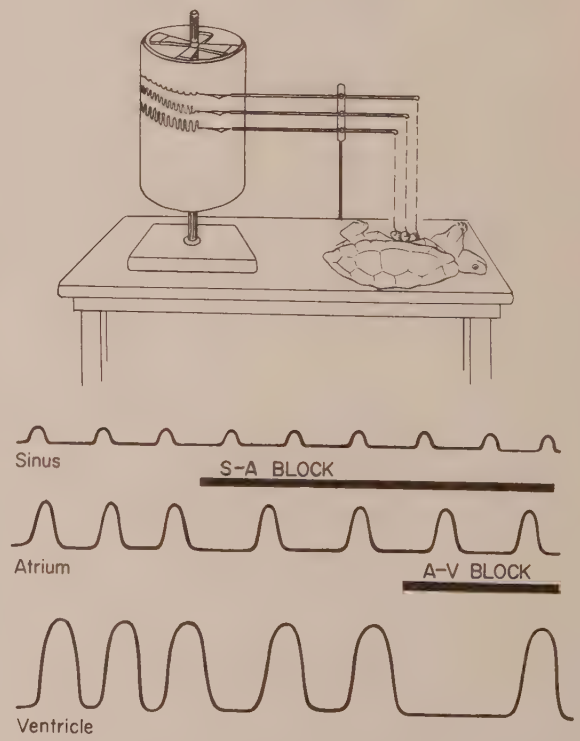


Figure 97. Method for recording the separate contractions of the sinus, atrium, and ventricle of the turtle heart. Below are illustrated the effects of S-A block and A-V block on impulse transmission.

Referring once again to Figure 97, it will be observed that the sinus contracts slightly ahead of the atrium and the atrium slightly ahead of the ventricle. Every time the sinus contracts, an electrochemical impulse spreads along the muscle fibers from the sinus to the atrium and then to the ventricle, making these parts of the heart contract in succession. Because the sinus initiates the contraction, it is called the *pacemaker* of the heart.

The human heart is different from that of the turtle, for no distinct sinus exists. However, located in the posterior wall of the right atrium immediately beneath the point of entry of the superior vena cava is a small area known as the *sino-atrial node* (S-A node) which is the embryologic remnant of the sinus found in lower animals. The rhythmic rate of contraction of muscle fibers excised from the S-A node is approximately 72 times per minute, while muscle excised from the atrium contracts about 60 times per minute and muscle from the ventricle about 25 times per minute. Because the S-A node has a faster rate of rhythm than other portions of the heart, impulses originating in the S-A node spread into the atria and ventricles, stimulating these areas so rapidly that they can never slow down to their natural rates of rhythm. As a result, the rhythm of the S-A node becomes the rhythm of the entire heart, and the S-A node is called the *pacemaker* of the heart.

Conduction of the Impulse Through the Heart

The heart is composed of two muscle latticeworks such as those shown in Figure 98. The two atria comprise one of the latticeworks and the two ventricles the other. An action potential initiated in any single fiber spreads over the membranes of all the fibers in each respective latticework. The two latticeworks in turn are joined together by the Purkinje system, which conducts impulses from the atrial latticework into the ventricular latticework. Thus, once a cardiac impulse begins in the S-A node it travels throughout the entire heart.

The Purkinje system. Even though the cardiac impulse can travel perfectly well along cardiac muscle fibers, a special conducting medium known as the *Purkinje system* performs two specific functions in conducting the cardiac impulse: First, it conducts the impulse from the atrial muscle latticework into the ventricular latticework; second, it increases the velocity of conduction in the ventricles. This system is composed of modified cardiac muscle fibers, called *Purkinje fibers,* that transmit impulses approximately six times as rapidly as normal heart muscle.



Figure 98. Cardiac muscle, showing the lattice arrangement of the fibers.

These fibers, illustrated in Figure 99, begin in the *atrioventricular node* (A-V node), which is located posteriorly in the lower part of the right atrial wall, and extend through the *A-V bundle* into the ventricular septum where they divide into two major bundles, one spreading along the wall of the right ventricle and the other along the wall of the left ventricle.

One of the major functions of the Purkinje system is to transmit the cardiac impulse throughout the ventricles as rapidly as possible, causing all portions to contract as nearly simultaneously as possible so that they will exert a coordinated pumping effort. Were it not for the Purkinje system, the impulse would travel much more slowly through the ventricles, allowing some of the muscle fibers to contract and relax before others. Obviously, this would result in decreased com-

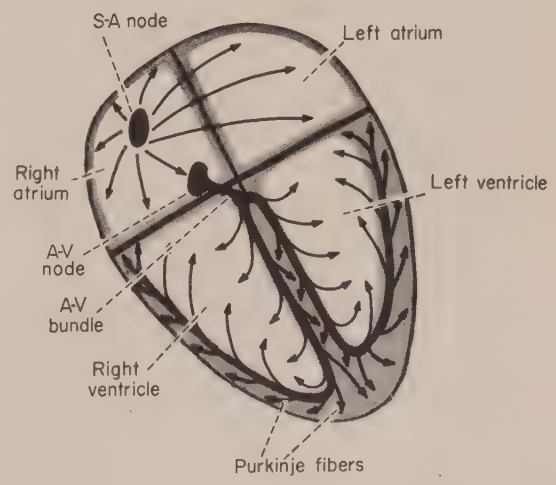


Figure 99. Transmission of the cardiac impulse from the S-A node into the atria, then into the A-V node, and finally through the Purkinje system to all parts of the ventricles.

pression of the blood and, therefore, decreased pumping power.

Sequence of impulse transmission through the heart—impulse delay at the A-V node. After the cardiac impulse originates in the S-A node, it travels first throughout the atria, causing them to contract. A few hundredths of a second after leaving the S-A node it reaches the A-V node and there enters the Purkinje system. However, the A-V node delays the impulse another few hundredths of a second before allowing it to pass on into the ventricles. This delay allows time for the atria to force blood into the ventricles prior to ventricular contraction. After the delay, the impulse spreads very rapidly through the Purkinje system, causing both ventricles to contract in full force within the next few hundredths of a second.

The A-V node delays the cardiac impulse by the following mechanism: The fibers in this node are quite different from either the usual cardiac muscle fibers or the Purkinje fibers, for they conduct impulses extremely slowly, at a velocity only one-tenth that of the cardiac muscle and only one-fiftieth that of the Purkinje fibers. Therefore, the cardiac impulse travels so slowly through the A-V node to the Purkinje system that a delay of more than one-tenth second occurs between contraction of the atrium and the ventricle.

IMPORTANCE OF THE A-V BUNDLE. Impulses cannot be conducted directly into the ventricular muscle from the atrial muscle in any way except through the A-V bundle. At all other points the two muscle masses are "insulated" from each other by fibrous septa. Therefore, malfunction of the A-V bundle can prevent normal transmission of impulses from the atria into the ventricles.

Block of impulse conduction through the heart. Occasionally the cardiac impulse is blocked at some point in its normal pathway because of damage to the heart. For instance, a portion of heart muscle or of the Purkinje system may be destroyed and replaced by fibrous tissue that cannot transmit the impulse. Figure 97 shows the effect of artificially blocking the impulse at two critical points in the turtle heart. First, a *sino-atrial block* has been effected by tying a string tightly around the heart between the sinus and the atrium. This stops impulse conduction from the sinus into the atrium, and, as is evident from the illustration, the sinus continues to beat at its own natural rate while the atrium and ventricle assume a rate equal to that of the natural rate of the atrium. In other words, the atrium becomes the pacemaker for the ventricle because the ventricle's natural rate of rhythmicity is much slower than that of the atrium. Another ligature is then tied tightly around the heart, this time between the atrium and the ventricle to cause *atrioventricular block.* After this, the sinus, the atrium, and the ventricle all beat at their own respective natural rates of rhythm. This experiment shows that the portion of the heart that beats most rapidly controls the rate of rhythm of the remainder of the heart only so long as functioning conductive fibers exist between the different areas.

In the human heart a block rarely occurs between the S-A node and the atrial muscle, but very frequently conduction from the atria into the ventricles through the A-V bundle is blocked. It is only through this bundle that the normal impulse can pass from the atria into the ventricles, because elsewhere the atria are connected to the ventricles not by conductive fibers but instead by fibrous tissue that cannot conduct impulses. Therefore, whenever the A-V bundle is blocked, the atria will beat at the rhythm of the S-A node, and the ventricles at their own natural rate. In other words, the rate of the atria will remain at approximately 72 beats per minute, while that of the ventricles will decrease to about 25 beats per minute. Despite this asynchrony of the atria and ventricles, the heart still operates reasonably satisfactorily as a pump, though its pumping ability may be decreased as much as 30 per cent. Nevertheless, it is evident that the atria are not absolutely essential for the heart to pump blood through the circulatory system.

Cessation of the impulse at the end of each heart beat—the refractory period. Normally, when an impulse spreads along the membranes of the heart muscle fibers, a second impulse cannot spread along these same membranes until approximately 0.3 second later, during which time the heart is said to be *refractory.* This period of inexcitability may be explained as follows: The normal muscle fiber has more negative than positive ions inside the membrane, and the cardiac impulse is caused by sudden transfer of some of these ions through the membrane so that more positive than negative ions now appear on the inside. Once this has occurred, another cardiac impulse cannot be transmitted until the physical state of the membrane changes back to its original state, as will be explained in Chapter 20; the time required for this to occur is about 0.3 second. Therefore, a second impulse cannot be conducted for this period of time.

Each impulse spreads through the ventricles of the normal heart in about 0.06 second, which is long before the fibers can conduct again. As a result, the impulse completely stops. The upper part of Figure 100 illustrates this principle, show-

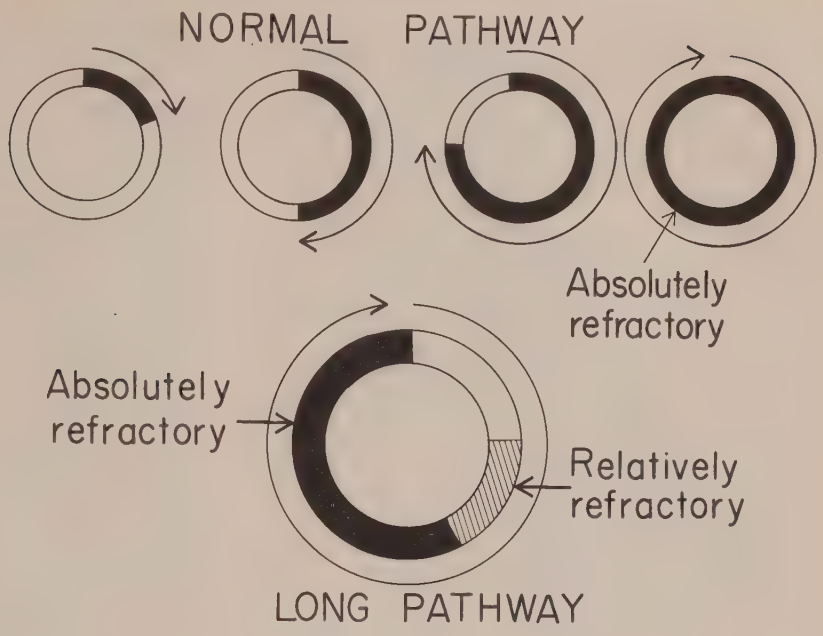


Figure 100. The principle of the circus movement.

ing a circular strip of cardiac muscle, in which an impulse starts at the 12 o'clock point, travels around the heart, and finally returns to the 12 o'clock point. When the impulse reaches the starting point the entire heart is still refractory, which causes the impulse to die.

The circus movement. Occasionally conditions become sufficiently abnormal for a cardiac impulse to continue on and on around the heart, never stopping. For instance, in the lower part of Figure 100, the length of the pathway has been increased so greatly by the larger circle (which represents an enlarged heart) that after traveling all the way around the heart the impulse now returns to the 12 o'clock position more than 0.3 second after it starts. By this time, the originally stimulated portion of the muscle is no longer refractory, which allows the impulse to travel around again. As the impulse proceeds, the refractory state recedes ahead of it, allowing it to continue indefinitely. This effect is known as a *circus movement.*

CAUSES OF CIRCUS MOVEMENTS. Circus movements can result from any of four different abnormalities of the heart. First, as explained in the preceding example, a circus movement is likely to occur when the heart becomes enlarged, thus creating a long pathway. A second cause is slow conduction of the impulse through the heart. For instance, failure of the Purkinje system causes the impulse then to be transmitted by the cardiac muscle itself. This slows impulse transmission about six-fold, often causing the impulse to return to the starting point after the originally excited muscle is no longer refractory. A third cause of a circus movement may be a decreased refractory period of the heart muscle. This sometimes results from increased cardiac excitability caused by epinephrine, sympathetic stimulation, or irritation of the heart as a result of disease. Fourth, and a very common cause, is transmission of impulses in figure 8's, in zigzags, or in any other odd pattern, the impulses sometimes traveling deep in the muscle and then later traveling at shallow levels, recrossing the same area that had already been stimulated at a deeper level. In this way the pathway becomes extremely lengthened, thus allowing development of an odd-shaped circus movement. Such a pathway, as illustrated in Figure 101B, causes ventricular fibrillation, which is discussed below.

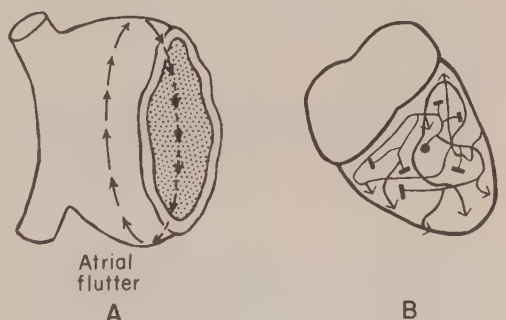


Figure 101. (A) Impulse transmission in *atrial flutter.* (B) Impulse transmission in *ventricular fibrillation.*

In any one of these instances it is possible for the heart muscle to recover more rapidly from the refractory state than usual. Then, when the impulse travels around the heart and returns to the starting point, the originally excited muscle will not be refractory and a circus movement will ensue.

EFFECT ON CARDIAC PUMPING. A circus movement is disastrous to the pumping action of the heart because normal pumping requires that the muscle relax as well as contract. A sufficiently long period of relaxation is not possible if the impulse continually travels through the muscular mass. Even when relaxation does occur, it occurs in only a portion of the heart at any one time. Therefore, the entire muscle mass never relaxes or contracts simultaneously, never allowing an alternating filling and squeezing action.

ATRIAL FLUTTER AND FIBRILLATION. Occasionally a circus movement occurs around and around the two atria, as shown in Figure 101A, at a rate of 200 to 400 times per minute, but not involving the ventricles. This causes *atrial flutter,* during which the atria seem to be "fluttering" rapidly but pumping almost no blood. At other times, a circus movement occurs in the atria at such a rapid rate with an odd-shaped pattern that one can see only minute fibrillatory movements of the muscle. This is called *atrial fibrillation,* and when it occurs the atria are of no use whatsoever as primer pumps for the ventricles.

VENTRICULAR FIBRILLATION. A circus movement having the very odd pattern illustrated in Figure 101B frequently develops in the ventricles, causing *ventricular fibrillation,* in which impulses travel in all directions. They go around refractory areas of muscle, dividing into two wavefronts in doing so. Some of the impulses die, but for every impulse that does so, another impulse divides to form two impulses, going in separate directions in the heart. This type of circus movement has been called the *chain-reaction movement,* for it is similar to the chain reaction that occurs in nuclear bomb explosions.

Ventricular fibrillation causes the ventricles to contract continuously in very fine, rippling fibrillatory movements. The contracting areas are widespread in the ventricles, interspersed with relaxing areas. As a consequence, the ventricles are incapable of pumping any blood, and the person dies immediately.

Ventricular fibrillation is frequently initiated by electric shock, particularly by 60 cycle alternating current; this causes impulses to go in many different directions at once in the heart, setting up the odd-shaped pattern of impulse transmission illustrated in Figure 101B. Ventricular fibrillation also occurs in diseased ventricles that become overly excitable, that develop a damaged Purkinje system, or that become greatly enlarged.

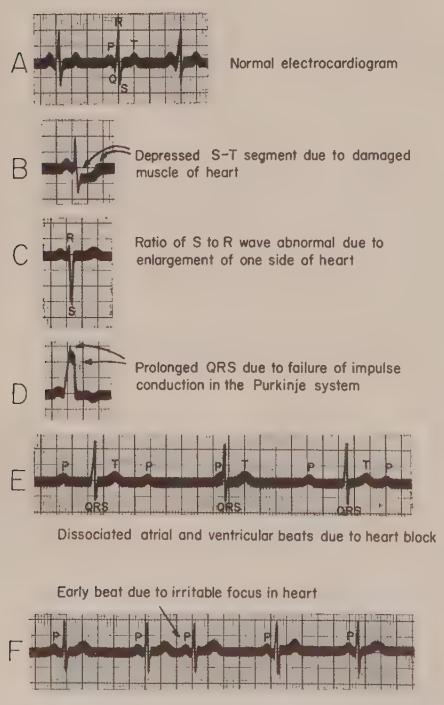


Figure 102. The normal and several abnormal electrocardiograms.

The Electrocardiogram

The electrocardiogram is a very important tool for assessing the ability of the heart to transmit the cardiac impulse. When an impulse travels over the muscle fibers, electrical current generated by the flowing ions spreads into the fluids surrounding the heart, and a minute portion of the current actually flows as far as the surface of the body. By placing electrodes on the skin over the heart or on any two sides of the heart and connecting these to an appropriate recording instrument, the impulse generated during each heart beat can be recorded.

In the normal electrocardiogram illustrated in Figure 102A, the curve labeled "P" is caused by electrical current spreading to the surface of the body when the impulse passes through the atria. The curves marked "Q," "R," and "S" are caused by passage of the impulse through the ventricles, and the curve "T" is caused by the return of ions into the ventricular muscle fiber at the end of the refractory period.

When cardiac abnormalities are caused by various diseases, the electrocardiogram often

becomes changed from the normal. Figure 102B shows that if some of the ventricular muscle has been damaged, the portion of the electrocardiogram between the S and T waves is depressed. This results from abnormal leakage of electrical current from the heart between heart beats. In general, it indicates damage to the membranes of the ventricular muscle fibers, which often occurs when one has an acute heart attack.

Figure 102C illustrates an abnormal QRS wave with excessive enlargement of the S wave and diminishment of the R wave. This is often caused by enlargement of one side of the heart, which makes more current flow from that side than from the other. High blood pressure very frequently provides this type of electrocardiogram.

Figure 102D shows what one finds in a person who has a partially blocked Purkinje system. In this instance, the impulse is transmitted through much of the ventricles so slowly that the QRS complex lasts for a prolonged period of time and also develops an abnormal shape.

Figure 102E illustrates the effect of blockage of the impulse at the A-V bundle. The P waves occur regularly, and the QRST waves also occur regularly but in no definite relationship to the P waves. The atria are beating at the natural rate of rhythm of the S-A node, while the ventricles have assumed their own rate of rhythm.

Finally, Figure 102F illustrates a record, indicated by the arrow, of a *premature beat* of the heart. The only abnormality here is that the impulse occurs too soon after the previous heart beat. This is usually caused by an irritable heart resulting from such factors as too much smoking, too much coffee drinking, or lack of sleep.

A study of Figure 102 shows how various abnormalities of heart function can be discovered from electrocardiographic recordings, and why this diagnostic procedure is used in all persons with heart disease.

FUNCTION OF THE HEART VALVES

Referring once again to Figure 96, one can see that the four valves of the heart are all oriented so that blood never flows backward but always forward when the heart contracts. The tricuspid valve prevents backflow from the right ventricle into the right atrium, the mitral valve prevents backflow from the left ventricle into the left atrium, and the pulmonary and aortic valves prevent backflow respectively into the right and left ventricles from the pulmonary and systemic arterial systems. These valves have the same functions as valves in any compression pump, for no pump of this type can possibly operate if fluid is allowed to flow backward during the compression cycle.

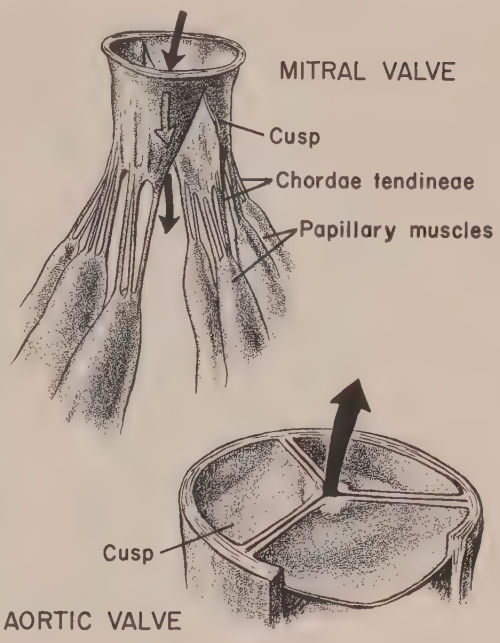


Figure 103. Structure of the mitral and aortic valves.

Figure 103 illustrates in more detail the structure of the valves. The tricuspid and mitral valves (the atrioventricular valves) are similar to each other, having rather expansive film-like vanes which are held in place by special ligaments, the *chordae tendineae*, extending from the papillary muscles. The pulmonary and aortic valves (the semilunar valves) are also similar to each other, having small, half-moon–shaped leaflets with very strong structures. The probable reason for these differences in the two types of valves is that blood must flow with great ease from the atria into the ventricles because the atria do not pump with much force. This requires very easily movable valves. The semilunar valves do not have to function with such extreme ease because of the great force of contraction of the ventricles; this allows the valves to be of simpler construction than the A-V valves.

The Heart Sounds

When one listens with a stethoscope to the beating heart he normally hears two sounds that are aptly described as "lub dub, lub dub, lub dub." The "lub" is called the *first heart sound* and the "dub" the *second heart sound.* The first sound is caused by closure of the A-V valves when

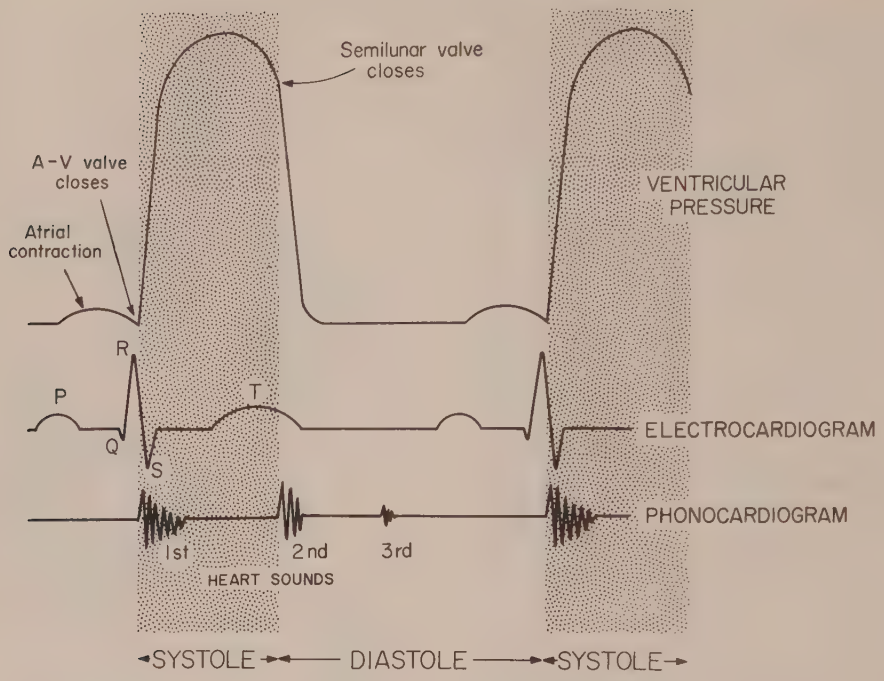


Figure 104. Relationship of ventricular pressure to the electrocardiogram and phonocardiogram during the cardiac cycle, showing also the periods of systole and diastole.

the ventricles first contract. This relationship to the cycle of heart beat is illustrated in Figure 104 by a *phonocardiogram,* which is a graphic representation of the heart sounds. When the ventricles contract, the increasing pressures in the two ventricles force the vanes of the A-V valves closed. The sudden stoppage of backflow from the ventricles into the atria creates vibration of the blood and heart walls, and the vibrations are transmitted through the chest to be heard as the first sound, the "lub" sound. Immediately after the ventricles have discharged their blood into the arterial systems, ventricular relaxation allows blood to begin flowing backward from the arteries into the ventricles, thereby causing the semilunar valves to close suddenly. This also sets up vibrations, this time in the blood of both the arteries and ventricles and also in the walls of the vessels and heart. These vibrations are transmitted to the chest causing the "dub" sound.

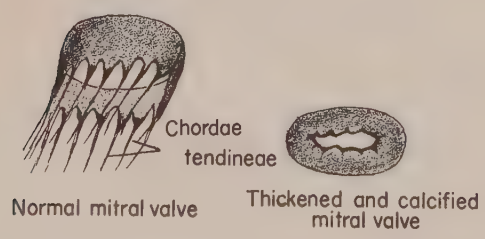


Figure 105. Effect of rheumatic fever on the mitral valve.

Valvular Heart Disease

The most frequent cause of valvular heart disease is rheumatic fever, a disease that results from an immune reaction to toxin secreted by streptococcic bacteria. The acute phase of the disease usually occurs two to three weeks after a person has had a streptococcic sore throat, scarlet fever, a streptococcic ear infection, or some other streptococcic infection. Antibodies formed against the toxin attack the valves, causing small, cauliflower-like growths on their edges which, over a period of several months to several years, erode the valves and cause ingrowth of fibrous tissue. The valve sometimes is eaten away completely or becomes so constricted and hardened that it cannot close in a normal manner. Figure 105 illustrates to the left a normal mitral valve and to the right a thickened and calcified valve resulting from several years of the rheumatic fever process. Obviously, such a valve is merely a constricted hole rather than a valve at all.

VALVULAR STENOSIS. All degrees of valvular destruction can develop following rheumatic fever. Sometimes the valve openings are so greatly narrowed by scar tissue that blood flows through

the opening only with much difficulty. This is called *stenosis.* If the aortic valve becomes stenosed, blood dams up in the left ventricle. If the mitral valve becomes stenosed, blood dams up in the left atrium and lungs. Likewise, stenosis of the pulmonary and tricuspid valves causes blood to dam up in the right ventricle and systemic circulation respectively.

VALVULAR REGURGITATION. Equally as often the valves do not become stenotic but instead are eroded so much that they cannot close. As a result, blood leaks backward through the valve that should be stopping this backward flow; this effect is called *regurgitation.* For instance, in aortic regurgitation much of the blood pumped into the aorta by the left ventricle returns to the ventricle during diastole rather than flowing on through the systemic circulation. Likewise, failure of the mitral valve to close when the left ventricle contracts allows blood to flow backward into the left atrium as well as forward into the aorta. Thus, leaking valves are as disastrous to cardiac function as are narrowed valvular openings. Occasionally a valve is both leaky and narrowed, decreasing the efficiency of cardiac pumping in two ways.

PULMONARY EDEMA. The valve most frequently affected by rheumatic fever is the mitral valve, though the aortic valve is affected almost as often. On the other hand, the valves of the right heart are rarely damaged severely. Severe damage to the mitral valve leads to excessive damming of blood in the left atrium and lungs, and in aortic valvular disease blood dams up in the left ventricle as well as in the left atrium and lungs. Therefore, following damage of either of these valves, blood engorges the lungs, causing fluid leakage from the pulmonary capillaries into the pulmonary tissues and alveoli, resulting in severe *pulmonary edema* and often drowning the patient in his own fluids. This will be explained in detail in Chapter 16 in the discussion of heart failure.

THE CARDIAC CYCLE

Now that the pumping action of the heart, the rhythmicity of the heart, and the function of the heart valves have been described, it is possible to synthesize this information into a sequence of events called the *cardiac cycle.* The cardiac cycle rightfully begins with initiation of the rhythmic impulse in the S-A node. Then transmission of the impulse through the heart causes the muscle fibers to contract. Thus, as shown in Figure 104, the P wave of the electrocardiogram occurs immediately before the pressure wave caused by atrial contraction. Approximately 0.15 second after the P wave begins, the electrical impulse has completed its passage through the atria and begins to spread rapidly over the ventricles, causing the QRS wave of the electrocardiogram and stimulating the ventricular muscle to contract. The rising ventricular pressure closes the A-V valves, thereby generating the first heart sound, and it opens the semilunar valves. The ventricles remain contracted approximately 0.3 second—that is, as long as the muscle remains refractory—and then they relax. During the process of relaxation, ions retransfer through the fiber membranes to reestablish the normal negative electrical charge inside the fibers. This causes the T wave of the electrocardiogram. Immediately after the ventricular muscle relaxes, a small amount of blood flows backward from the arteries into the ventricles, closing the semilunar valves which elicits the second heart sound. Following ventricular relaxation, no further contraction occurs until a new electrical impulse is initiated in the S-A node.

Systole and Diastole

The period during the cardiac cycle when the ventricles are contracting is called *systole,* and the period of relaxation is called *diastole.* A clinician examining the heart can note the periods of systole and diastole either from the electrocardiogram or from the heart sounds; systole begins with the QRS wave and ends with the T wave, or it begins with the first heart sound and ends with the second sound. Diastole on the other hand, begins with the T wave and ends with the QRS wave, or it begins with the second heart sound and ends with the first heart sound.

Sometimes it is quite important to distinguish between systole and diastole. This is particularly true when one is studying valvular disorders or abnormal openings between the two sides of the heart. For instance, leakage of a semilunar valve causes a "swishing" sound called a *murmur* during diastole, because that is the period of the cardiac cycle when abnormal semilunar valves leak instead of closing tightly. On the other hand, a murmur caused by leakage of an A-V valve occurs during systole, because that is the period when these valves leak if they are abnormal.

Pressure Changes during the Cardiac Cycle

Figure 106 shows the changes in pressure in the left atrium, left ventricle, and aorta during a typical cardiac cycle. During diastole, the left atrial pressure is a little higher than that of the

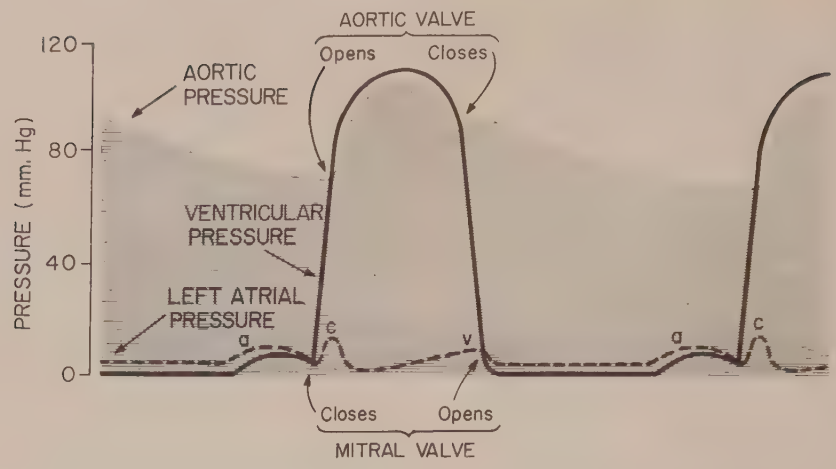


Figure 106. Pressures in the aorta, left ventricle, and left atrium during the cardiac cycle.

left ventricle. This is the reason blood flows continually from the left atrium into the left ventricle. Toward the end of diastole, contraction of the left atrium elevates the left atrial pressure to an even higher value and forces an extra quantity of blood into the ventricle. Then, suddenly, the left ventricle contracts, the A-V valves close, and the ventricular pressure rises rapidly. When the pressure rises higher than that in the aorta, the aortic valve opens and allows blood to flow into the aorta during the entire remainder of systole. Immediately after the ventricle relaxes, the ventricular pressure falls precipitously, allowing a slight backflow of blood that immediately closes the aortic valve. Throughout diastole the aortic pressure remains high because a large quantity of blood has been stored in the very distensible arteries during systole. This blood runs off slowly through the capillaries back to the veins, allowing the aortic pressure to fall from a peak during systole of approximately 120 mm. Hg down to a minimum of approximately 80 mm. Hg by the end of diastole. Therefore, the normal blood pressure is said to be 120/80, meaning by this 120 mm. Hg *systolic pressure* and 80 mm. Hg *diastolic pressure.*

THE LAW OF THE HEART

The amount of blood pumped by the heart is normally determined by the amount of blood flowing from the veins into the right atrium. This principle is frequently called the "law of the heart." That is, the heart is simply an automaton that continues to pump all of the time, and whenever blood enters the right atrium it is pumped on through the heart. Of course, there is a maximum rate at which the heart can pump, for which reason, to be completely accurate, the law of the heart is more correctly stated as follows: *Within physiologic limits, the heart pumps all of the blood that flows into it without excessive damming of blood in the veins.* Thus, the heart is analogous to a sump pump, because any time fluid enters the chamber of a sump pump, it is pumped out of the chamber immediately. In the case of the heart, whenever blood enters one of the atria, it is immediately pumped on into a ventricle and then on into the arteries.

It is a special quality of cardiac muscle that gives the heart this ability to pump varying amounts of blood depending upon the rate of venous inflow. When cardiac muscle is stretched prior to contraction, it contracts with greater force than when it is not preliminarily stretched. When only a small quantity of blood enters the heart, the muscle fibers are not stretched greatly, and the force of contraction is weak. On the other hand, if large quantities of blood enter, the heart chambers dilate greatly, the muscle fibers stretch, and the force of contraction becomes very great. As a result, the increased quantity of blood returning to the heart is pumped on through as a result of the intrinsic adaptation of the heart.

The law of the heart is often inapplicable after the heart has been damaged, for then even normal quantities of blood returning to the heart cannot be pumped with ease. As a consequence, blood begins to dam up in the veins of either the lungs or the systemic circulatory system. In this case the heart is said to be *failing.* The subject of heart failure is so very important that it will be discussed in detail in Chapter 16.

The heart-lung preparation. One of the methods frequently used in the physiology laboratory to demonstrate the independent ability of the heart to pump blood is the heart-lung preparation illustrated in Figure 107. In this preparation, blood flow through the systemic circulation is stopped, but it is simultaneously channeled from the aorta through an external system and then back into the right atrium. Artificial respiration is supplied to the lungs to keep the blood appropriately oxygenated, and nutrition in the form of glucose is occasionally added to the blood. The heart will continue to beat for many hours with the external flow system taking the place of the systemic circulation. One can vary the resistance of the external circuit by tightening or loosening a screw clamp. He can measure the blood flow with a flowmeter, and can change the amount of blood flowing into the left atrium by raising or lowering the venous reservoir.

Several principles that demonstrate the independent function of the heart can be illustrated beautifully with the heart-lung preparation. First, it can be shown that within physiologic limits raising or lowering the venous reservoir is the main factor that determines how much blood will be pumped. In other words, *the greater the input pressure forcing blood from the veins into the heart, the greater the volume of blood pumped,* which is the "law of the heart."

Second, it can be demonstrated that changing the resistance of the circuit within normal limits does not greatly affect the amount of blood pumped. However, the pressure in the aorta increases approximately in proportion to the increase in resistance. That is, *the amount of resistance to blood flow through the circulatory system changes arterial pressure greatly, but hardly affects the amount of blood pumped by the heart* unless the resistance becomes so great that the heart simply cannot pump with enough force to overcome it.

A third interesting effect that can be demonstrated is that of changing the elasticity of the arteries. This can be done by connecting an air bottle to the arterial system. Blood flows into the bottle and compresses the air when the pressure rises during systole, but during diastole the compressed air forces blood back into the system, thereby helping to maintain a relatively even level of pressure. With the bottle in the system, the arterial pressure pulsates far less than when it is out of the system.

Finally, the heart-lung preparation can be used to demonstrate the effect of temperature, drugs, or abnormal blood constituents on the heart. For instance, *increasing the temperature of the blood 10° F. increases the heart rate approximately 100 per cent.* Also, increasing the amount of *calcium* in the blood makes the heart contract with increased vigor, while increasing the amount of *potassium* decreases the vigor of heart contraction. Finally, when the nutrients of the blood fall to low values, addition of glucose or other nutrients can greatly stimulate the function of the heart.

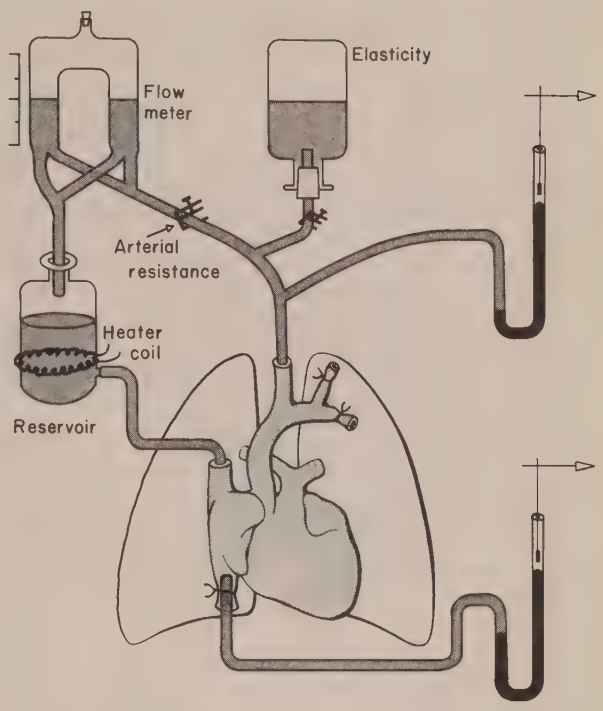


Figure 107. The heart-lung preparation.

NERVOUS CONTROL OF THE HEART

Though the heart has its own intrinsic control systems and can continue to operate without any nervous influences, there are times when efficacy of heart action can be changed greatly by regulatory impulses from the central nervous system. The nervous system is connected with the heart through two different sets of nerves, the *parasympathetic nerves* and the *sympathetic nerves.* The connections of the parasympathetic (vagi) and sympathetic nerves with the heart are shown in Figure 108.

Stimulation of the parasympathetic nerves causes the following four effects on the heart: (1) decreased rate of rhythmicity of the S-A node, (2) decreased force of contraction of the atrial muscle, (3) decreased rate of conduction of impulses through the A-V node, which lengthens

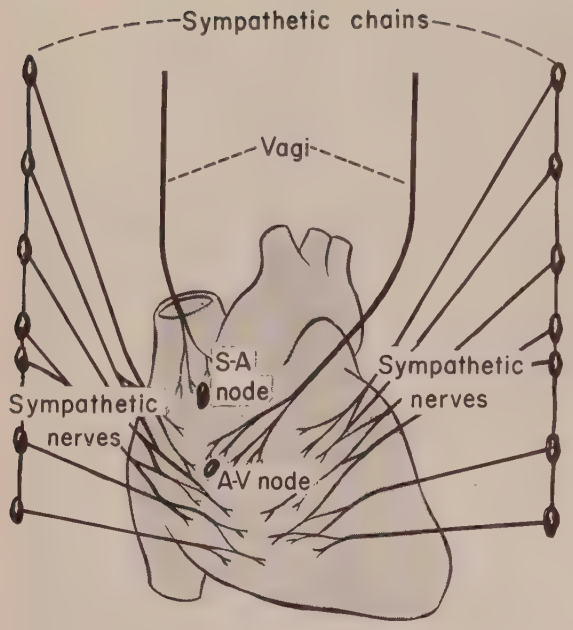


Figure 108. Innervation of the heart.

the delay period between atrial and ventricular contraction, and (4) decreased blood flow through the coronary blood vessels which are the vessels that supply nutrition to the heart muscle itself. All of these effects may be summarized by saying that parasympathetic stimulation decreases all activities of the heart. Usually, the heart is stimulated by the parasympathetics during periods of rest; this allows the heart to rest at the same time the remainder of the body is resting. This perhaps preserves the resources of the heart, for without periods of rest the heart probably would wear out at a much earlier age than it normally does.

Stimulation of the sympathetic nerves has essentially the opposite effects on the heart: (1) increased heart rate, (2) increased vigor of cardiac contraction, and (3) increased blood flow through the coronary blood vessels to supply increased nutrition to the heart muscle. These effects can also be summarized by saying that sympathetic stimulation increases the activity of the heart as a pump. This effect is necessary when a person is subjected to stressful situations such as exercise, disease, excessive heat, and other conditions that demand rapid blood flow through the circulatory system. Therefore, the sympathetic effects on the heart are a stand-by mechanism held in readiness to make the heart beat with extreme vigor when necessary.

REFERENCES

Clowes, G. H. A., Jr.: Extracorporeal maintenance of circulation and respiration. *Physiol. Rev.* 40:826, 1960.

Cranefield, P. F., and Hoffman, B. F.: Electrophysiology of single cardiac cells. *Physiol. Rev.* 38:41, 1958.

Hoffman, B. F., de Carvalho, A. P., Mello, W. C., and Cranefield, P. F.: Electrical activity of single fibers of the atrioventricular node. *Circ. Res.* 7:11, 1959.

Hoffman, B. F., and Cranefield, P. F.: *Electrophysiology of the Heart.* New York, Blakiston, 1960.

Hurst, J. W.: *Cardiac Resuscitation.* Springfield, Ill., Charles C Thomas, 1960.

Scher, A. M., Rodriquez, M. I., Liikane, J., and Young, A. C.: The mechanism of atrioventricular conduction. *Circ. Res.* 7:54, 1959.

Warner, H. R., and Toronto, A. F.: Regulation of cardiac output through stroke volume. *Circ. Res.* 8:549, 1960.

Whitehorn, W. V.: Heart. *Ann. Rev. Physiol.* 25:235, 1963.

BLOOD FLOW THROUGH THE SYSTEMIC CIRCULATION AND ITS REGULATION

13

All of the circulation besides the heart and the pulmonary circulation is called the *systemic circulation.* The blood flowing through this part of the circulation provides nutrition to the tissues, transport of excreta from the tissues, cleansing of the blood as it passes through the kidneys, absorption of nutrients from the gastrointestinal tract, and mixing of all the fluids of the body, as explained in Chapter 1. The rate of blood flow to each respective tissue is almost exactly the amount required to provide adequate nutrition, no more, no less. The purpose of the present chapter, therefore, is to describe, first, the basic principles of blood flow through the circulation and, second, the mechanisms that control the blood flow to each respective tissue in proportion to its needs.

HEMODYNAMICS

The study of the physical principles that govern blood flow through the vessels and the heart is known as *hemodynamics.* The heart forces a stream of blood into the aorta, distending it and creating pressure within it. This pressure then pushes the blood through the arteries, arterioles,

capillaries, venules, and veins, finally back to the heart. As long as the animal remains alive, this flow of blood around the continuous circuit never ceases.

The small arteries, arterioles, capillaries, venules, and small veins have such small diameters that blood flows through them with considerable difficulty. In other words, the vessels are said to offer *resistance* to blood flow. Obviously, the smaller the vessel the greater is the resistance, and the larger the vessel the less the resistance.

In essence, the discussion in the present chapter centers around the effects of pressure and resistance on blood flow.

Blood Flow

The amount of blood pumped by the heart when a person is at rest is approximately 5 liters per minute. This can increase to as much as 20 to 30 liters per minute during the most extreme exercise, or it can decrease to as low as 1.5 liters per minute following severe hemorrhage.

Blood flow to the different parts of the body during rest are given in Table 2. One will note that the brain receives approximately 14 per cent of the total blood flow, the kidneys 22 per cent, the liver 27 per cent, and the muscles, which comprise over half of the body, only 15 per cent. However, during exercise quite a different picture develops, for then essentially all of the increase in blood flow occurs in the muscles, this increasing as much as 15-fold during very intense exercise and representing as much as 75 per cent of the total blood flow.

Table 2. *Blood Flow to Different Organs and Tissues Under Basal Conditions. (Based mainly on data compiled by Dr. L. A. Sapirstein.)*

	Per cent	*ml./min.*
Brain	14	700
Heart	3	150
Bronchi	3	150
Kidneys	22	1100
Liver	27	1350
Portal	(21)	(1050)
Arterial	(6)	(300)
Muscle	15	750
Bone	5	250
Skin (cool weather)	6	300
Thyroid gland	1	50
Adrenal glands	0.5	25
Other tissues	3.5	175
Total	100.0	5000

Measurement of blood flow. Several different types of apparatus have been designed to measure blood flow. Some measuring apparatus connect with a blood vessel in such a manner that the blood must flow through the apparatus, while others employ indirect methods that do not require opening the vessels.

Figure 109A illustrates the *rotameter* type of flowmeter which must be connected directly to the vascular circuit. Blood enters at the bottom and lifts the float in the conical tube. The greater the flow the higher the float will rise. A steel shaft connected to the float protrudes upward through a coil of wire, and this in turn is connected to an electronic measuring apparatus to record the rate of blood flow by detecting the change in magnetic properties of the coil.

A second type of flowmeter, the *electromagnetic flowmeter*, is shown in Figure 109B. The electromagnet creates a very strong magnetic field passing through the blood vessel. As the blood flows through the field, it "cuts" the magnetic lines of force and develops an electrical potential at right angles to these lines of force. Two electrodes placed at right angles to the magnet pick up this potential and transmit it into an appropriate electronic recording apparatus. The blood vessel does not have to be opened when this type of flowmeter is used, and the response of the instrument is so rapid that it can measure even very transient changes in flow that take place in as little as $\frac{1}{100}$ of a second. Because of these advantages, this type of flowmeter has become very widely used in physiological studies of the circulation.

THE PLETHYSMOGRAPH. Figure 110A illustrates the *plethysmograph*, which can be used to measure blood flow in a limb or in any other protruding portion of the body. The arm is inserted into the open end of the chamber, and an airtight seal is made between the end of the chamber and the forearm. The cuff around the upper arm is suddenly inflated to create a pressure higher than the pressure in the veins but lower than that in the arteries (about 40 mm. Hg). This allows blood to continue flowing into the arm through the arteries but stops all flow away from the arm through the veins. As a result, the arm swells, forcing air out of the chamber into a recording drum. One can tell how much the arm swells by injecting air into the chamber with a syringe until the record rises an amount equal to that caused by the swelling.

Figure 110B shows a typical recording from the plethysmograph, depicting a gradual increase in volume of the arm. A straight line drawn upward along the initial rise in the record illustrates that the initial rate at which the arm swells is approxi-

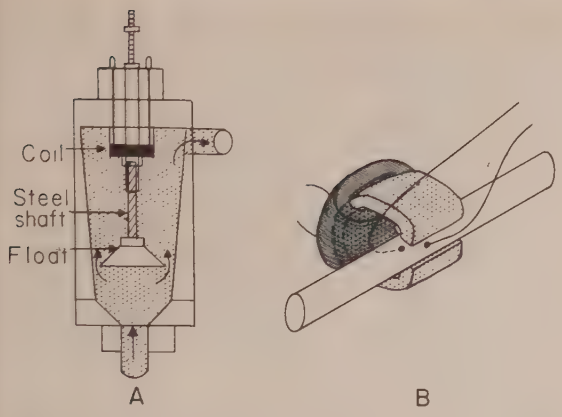


Figure 109. Instruments for recording blood flow: (A) the rotameter, (B) the electromagnetic flowmeter.

mately 3 ml. in 2 seconds. In other words, the *blood flow* into the arm is 3 ml. per 2 seconds or 90 ml. per minute.

Velocity of blood flow in different portions of the circulation. The term *blood flow* means the actual quantity of blood flowing through a vessel or group of vessels in a given period of time. In contradistinction to this, the *velocity* of blood flow means the distance that the blood travels along a vessel in a given period of time.

If the quantity of blood flowing through a vessel remains constant, the velocity of blood flow obviously decreases as the size of the vessel increases. The aorta as it leaves the heart has a cross-sectional area of approximately 2 sq. cm. Then it divides into a number of large arteries, a portion of the aortic blood flowing into each of these vessels. The total cross-sectional area of the branching arteries is considerably greater than that of the aorta. Therefore, the average velocity of blood flow decreases. Then the large arteries divide into still smaller arteries, and these into arterioles. The total cross-sectional area of the arterioles is 15 to 30 times that of the aorta, and the velocity of blood flow is proportionately less. The cross-sectional area of the capillaries is about 750 times that of the aorta, the venules about 60 times, and the veins about 4 times. It is evident then that the cross-sectional areas of the vessels leaving the heart and entering the heart are rather small, but these spread out like the limbs of a tree into a very broad cross-sectional area in the region of the capillaries. As a consequence, the velocity of blood flow is greatest in the aorta and least in the capillaries. Numerically, the velocities are approximately the following: aorta, 30 cm. per second; arterioles, 1.5 cm. per second; capillaries, 0.4 mm. per second; venules, 5 mm. per second; and venae cavae, 8 cm. per second.

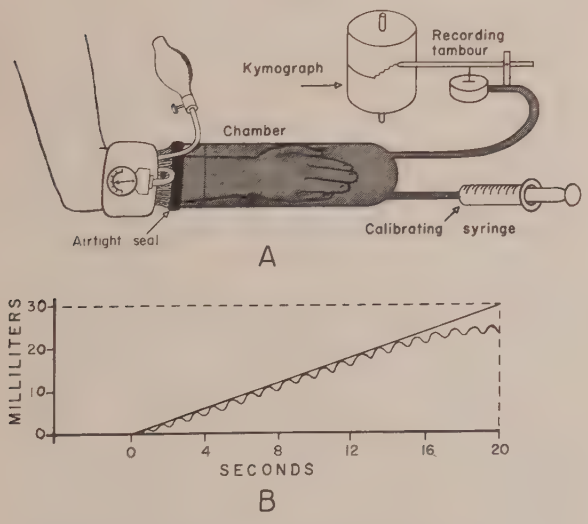


Figure 110. (A) The plethysmograph. (B) A typical recording from the plethysmograph.

TRANSIT TIME FOR BLOOD IN THE CAPILLARIES. The velocity of blood flow in the capillaries is particularly significant because it is here that oxygen, other nutrients, and excreta pass back and forth between the blood and the tissue spaces. The length of the average capillary is about 0.5 to 1 mm. Therefore, at a velocity of blood flow of 0.4 mm. per second, the length of time that blood remains in each capillary averages approximately one to two seconds. Despite this short time, an extremely large proportion of the substances in the blood can transfer through the capillary membrane. This signifies once again the extreme capability of the capillaries as an exchange mechanism between the blood and the tissue fluids. Without this exchange in the capillaries, all the other hemodynamic functions of the circulation would be useless.

Blood Pressure

The pressure in a blood vessel is the force that the blood exerts against the walls of the vessel. This force distends the vessel because all blood vessels are distensible, the veins 6 times as much so as the arteries. Pressure also makes blood attempt to leave a vessel by any available opening, which means that the normally high pressure in the arteries forces blood through the small arteries, then through the capillaries, and finally into the veins. The importance of blood pressure, then, is that it is the force that makes blood continue to flow through the circulation.

Measurement of blood pressure—the mercury manometer. The standard device for measuring blood pressure is the mercury manometer shown

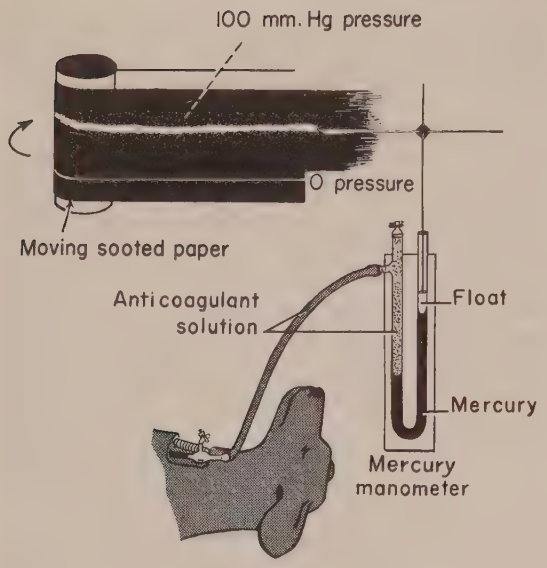


Figure 111. A mercury manometer.

in Figure 111. The rubber tube is attached to a blood vessel. The blood from the vessel presses downward on the left column of mercury forcing the right column upward. A float and recording arm resting on the mercury rise and fall with the level of the mercury and thereby record the pressure. If the level of mercury in the right column is 100 millimeters above the level in the left column, the blood pressure is said to be 100 millimeters of mercury (mm. Hg). If the level to the right is 200 millimeters above that on the left, then the pressure is 200 mm. Hg.

Because of the convenience of the mercury manometer for measuring pressure, almost all pressures of the circulatory system are expressed in millimeters of mercury. However, pressure can also be expressed as water pressure because one can equally as well connect an artery to a column of water and determine how high the water rises. Water weighs 13.5 times less than mercury, which means that the level of water in a manometer will rise this many times as high as the level of mercury in a manometer. Therefore, 13.5 mm. or 1.35 cm. of water pressure equals 1 mm. Hg pressure.

HIGH FIDELITY METHODS FOR RECORDING BLOOD PRESSURE. Unfortunately, the mercury in the mercury manometer has so much *inertia* that it cannot rise and fall rapidly. For this reason, this type of manometer, though excellent for recording mean pressure, cannot respond to pressure *changes* that occur more rapidly than one cycle every two to three seconds. Therefore, when it is desired to record rapidly changing pressures, some other type of pressure recorder such as one of those illustrated in Figure 112 is required. Each of the three transducers illustrated in this figure converts pressure into an electrical signal that is then recorded on a high-speed recorder. A syringe needle is connected to each transducer; this needle is inserted directly into the vessel whose pressure is to be measured, and the pressure is transmitted into a chamber bounded on one side by a thin membrane. In transducer A this membrane forms one plate of a variable capacitor. As the pressure rises, the membrane moves closer to the other plate and changes the electrical *capacitance,* which is detected and recorded by an electronic instrument. In transducer B a small iron slug on the membrane moves upward into a coil and changes the coil's *inductance,* which likewise is detected and recorded. In transducer C the membrane is connected to a thin, stretched wire. As the membrane moves, the length of the wire changes, which changes its *resistance,* once again allowing the pressure to be recorded by an electronic recorder.

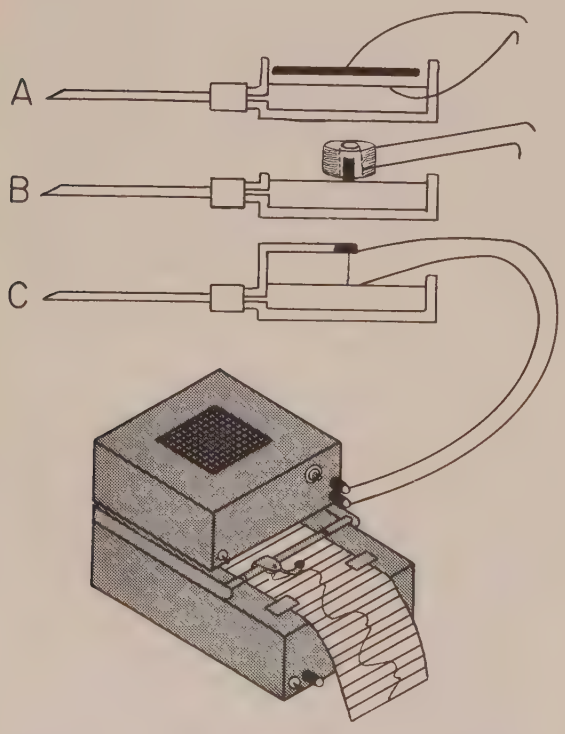


Figure 112. Principles of operation of three different electronic transducers for recording rapidly changing blood pressure.

Relationship of pressure to blood flow. When the blood pressure is high at one end of a vessel and low at the other end, blood will attempt to flow from the high toward the low pressure area because the two forces are unequal. The rate of blood flow is directly proportional to the difference between the pressures. Figure 113 illustrates

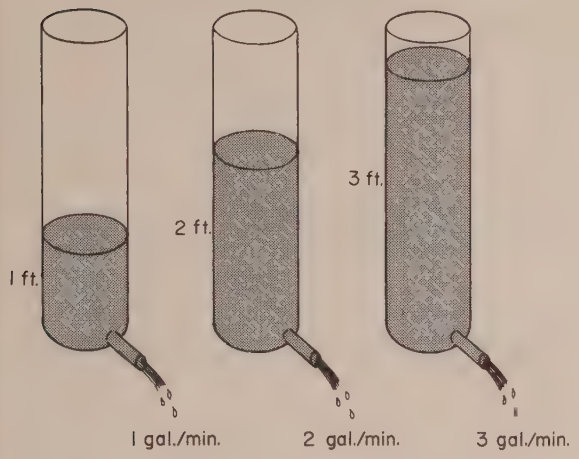


Figure 113. Effect of pressure on flow.

this principle, showing that the flow from the spout becomes greater in proportion to the water level in the chamber.

It is not the actual pressure in a vessel that determines how rapidly blood will flow, but the *difference* between the pressures at the two ends of the vessel. For instance, if the pressure at one end of the vessel is 100 mm. Hg and 0 mm. Hg at the other end, the pressure difference is 100 mm. Hg. If the pressures at the two ends of this same vessel are suddenly changed to 5000 mm. Hg at one end and 4950 mm. Hg at the other end, the pressure difference now is only 50 mm. Hg, and the flow will be reduced to one half as much provided the vessel's diameter does not become changed by the pressure.

Resistance to Blood Flow

Resistance is essentially the same as friction, for it is friction between the blood and the vessel walls that creates impediment to flow. The amount of resistance is dependent on the length of the vessel, the diameter of the vessel, and the viscosity of the blood.

Effect of vessel length on resistance to blood flow. The longer a vessel, the greater the vascular surface along which the blood must flow, and consequently the greater the friction between the blood and vessel wall. For this reason, the *resistance to blood flow is directly proportional to the length of the vessel.*

This principle is illustrated by the three tubes in Figure 114, each of which has a pressure head of 100 mm. Hg at its inlet. The tube with a length of 1 has an output of 3 ml. per minute while the one with a length of 3 but of exactly the same diameter has only one-third as much output, illustrating the inverse relationship between the length of the tube and the amount of fluid that will flow when all other factors remain constant.

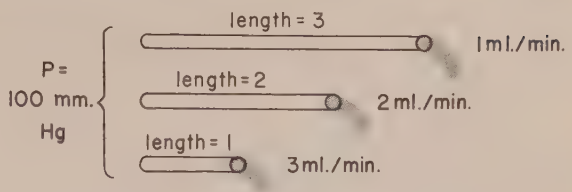


Figure 114. Effect of vessel length on flow.

Effect of vessel diameter on resistance to blood flow. Fluid flowing through a tube is retarded only along the walls. For this reason, the velocity of blood flow in the middle of a vessel is very great while the velocity along the surface is very low, and the larger the vessel the more rapidly can the central portion of blood flow. Because of this effect, the blood flow through a vessel—that is, the total quantity of blood that passes through the vessel each minute—increases very markedly as the diameter of the vessel increases. In fact, if all other factors stay constant, the blood flow through the vessel is directly proportional to the *fourth power* of the diameter. Thus, in Figure 115 three vessels of the same length but with diameters of 1, 2, and 4 are shown. Note that even these small changes in diameter increase the flow respectively 16- and 256-fold. Because of this extreme dependence of flow on diameter, very slight changes in vascular diameter can affect tremendously blood flow to different vascular regions. The diameter of most vessels can change approximately 4-fold, which can change the blood flow as much as 256-fold.

Effect of viscosity on blood flow. The more viscous the fluid attempting to flow through a tube the greater the friction with the wall, and, consequently, the greater the resistance. This effect is illustrated in Figure 116, which shows water, plasma, and normal blood attempting to go through spouts of equal dimensions. In each instance the amount of fluid flowing through the spout is 100 ml. per minute, but, because of different viscosities of the three fluids, the pressure

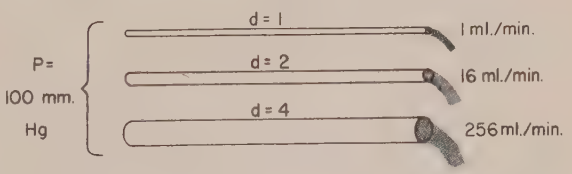


Figure 115. Effect of vessel diameter on flow.

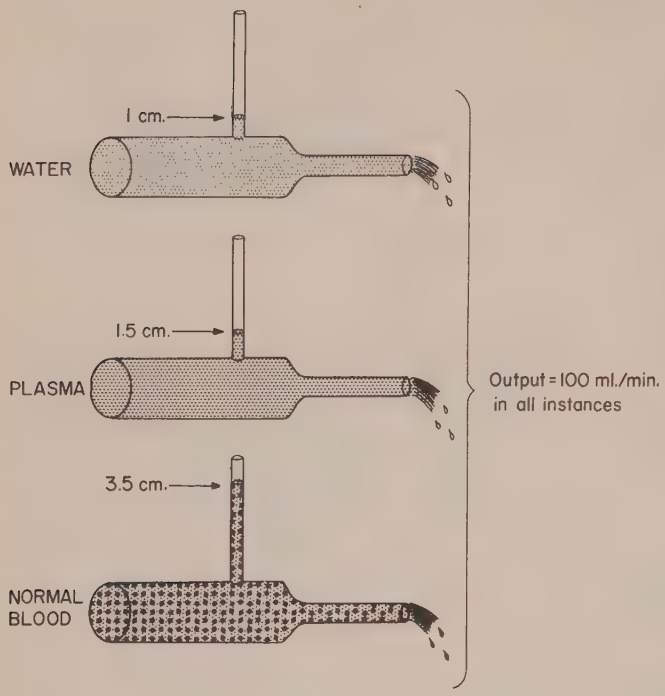


Figure 116. Effect of viscosity on flow.

required to force each through the spout is also different. The level of fluid in the vertical tube, which is an indicator of pressure, is 1 cm. when water is flowing, 1.5 cm. for plasma, and 3.5 cm. for blood. In other words, the relative viscosities of water, plasma, and blood are 1, 1.5, and 3.5.

The most important factor governing blood viscosity is the concentration of red blood cells. The viscosity of normal blood is about 3.5 times that of water. However, when the red blood cell concentration falls to one-half normal, the viscosity becomes only 2 times that of water, and when the concentration rises approximately 70 per cent above normal, the viscosity may rise to as high as 20 times that of water. For these reasons, the flow of blood through the blood vessels of anemic persons—that is, persons who have low concentrations of red blood cells—is extremely rapid, while blood flow is very sluggish in persons who have excess red blood cells (polycythemia).

Interrelationship of Pressure, Flow, and Resistance

It is obvious from the preceding discussions that pressure and resistance oppose each other in affecting blood flow. One can state this relationship mathematically by the following formula:

$$\text{Blood flow} = \frac{\text{Pressure}}{\text{Resistance}}$$

The same formula can be expressed in two other algebraic forms:

$$\text{Pressure} = \text{Blood flow} \times \text{Resistance}$$

$$\text{Resistance} = \frac{\text{Pressure}}{\text{Blood flow}}$$

These formulas are basic in almost all hemodynamic studies of the circulatory system, and therefore must be thoroughly understood prior to any attempt to analyze the operation of the circulation.

Poiseuille's law. By inserting the various factors that affect resistance into the preceding formulas, one can derive still another formula known as Poiseuille's law:

$$\text{Blood flow} = \frac{\text{Pressure} \times (\text{Diameter})^4}{\text{Length} \times \text{Viscosity}}$$

This formula expresses the ability of blood to flow through any given vessel, showing that the rate of blood flow is directly proportional to the pressure difference between the two ends of the vessel, directly proportional to the fourth power of the vessel diameter, and inversely proportional to the vessel length and blood viscosity.

REGULATION OF BLOOD FLOW THROUGH THE VESSELS

The Arterioles as the Major Regulators of Blood Flow

The rate of flow through each respective tissue is controlled mainly by the arterioles. There are three separate reasons for this:

First, approximately one half of all the resistance to blood flow in the systemic circulation is in the arterioles, which can be seen by studying Figure 117. Therefore, a slight change in arteriolar diameter can change the total resistance to flow through any given tissue far more than can a similar change in diameter in any other vessel.

Second, the arterioles have a very strong muscular wall constructed in such a manner that the diameter of an arteriole can change as much as 3- to 5-fold. Remembering once again that resistance is inversely proportional to the fourth power of a vessel's diameter, it immediately becomes obvious that the resistance to blood flow through the arterioles can be changed as much as several hundred- or even a thousand-fold by simply relaxing or contracting the smooth muscle walls.

Third, the smooth muscle walls of the arterioles respond to two different types of stimuli that regulate blood flow. First, they respond to the local needs of the tissues, increasing blood flow when the supply of nutrients to the tissues falls too low and decreasing the flow when the nutrient supply becomes too great. This mechanism is called *autoregulation.* Second, autonomic nerve impulses, particularly *sympathetic impulses,* have a profound effect on the degree of contraction of the arterioles.

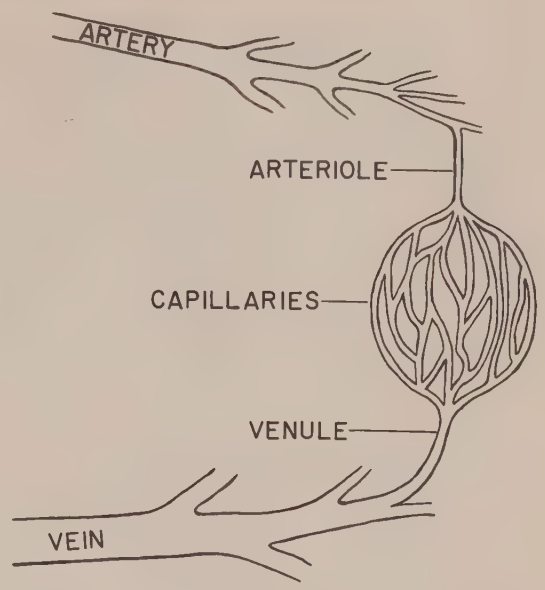


Figure 117. Demonstration of the minute size of the arteriole relative to the remainder of the vascular system, which explains its tremendous resistance.

Autoregulation—Relation to Tissue Oxygen Need

The term "autoregulation" means automatic adjustment of blood flow to each tissue in proportion to its need for nutrients. The most important nutrient that must be supplied to the tissues is *oxygen,* and recent experiments have shown that the oxygen concentration in the tissues is a very powerful stimulus for autoregulation. If, for instance, a tissue becomes more active than usual, the need for oxygen may increase as much as 5- to 10-fold, and the blood flow automatically increases as much as 3- to 6-fold to supply this oxygen.

Mechanism by which oxygen deficiency causes arteriolar dilatation. Even though we know that blood flow through a tissue is controlled in proportion to its need for nutrients, and especially in proportion to its need for oxygen, we still do not know the exact mechanism by which oxygen need or need of other nutrients can affect the degree of vascular contraction. There are two basic theories for arteriolar vasodilatation. One of these, the "oxygen need" theory, is illustrated in Figure 118.

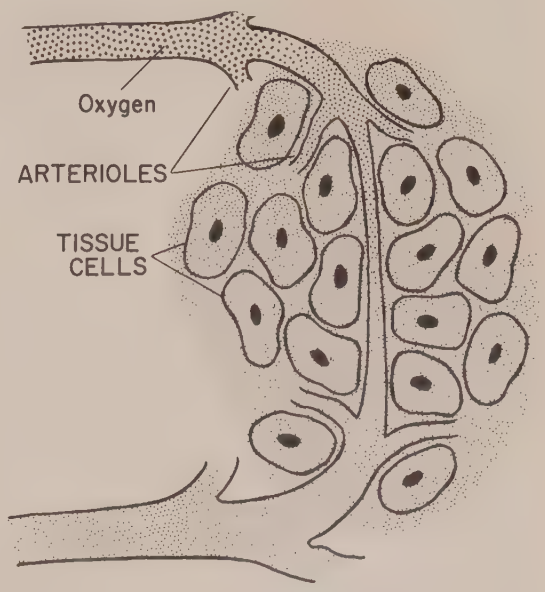


Figure 118. Postulated mechanism of autoregulation.

This shows that the tissue cells lie in close proximity to the arterioles and that they are in competition with the arterioles for the oxygen represented by the small dots. It is assumed that the smooth muscle of the arteriolar wall requires oxygen to contract. Should the oxygen supply diminish, the vessels would be expected to dilate simply because of failure to receive adequate oxygen to maintain contraction. Also, if the tissue utilizes an increased quantity of oxygen, the amount of oxygen available to the arterioles would decrease as well and cause increased blood flow. The resulting increase in blood flow then increases the tissue oxygen concentration back toward a normal level.

Another theory is that oxygen lack causes some *vasodilator substance* to be formed in the tissue. That is, decreased availability of oxygen to the metabolic systems of the cells supposedly causes them to secrete a chemical substance that acts directly on the blood vessels to cause active vasodilatation. Some of the suggested vasodilator substances have been histamine, lactic acid, carbon dioxide, and various breakdown products of adenosine triphosphate, such as adenine. Thus far, however, none of these substances has been proved to be secreted in sufficient quantity to account for more than a small fraction of the autoregulation response.

Long-term autoregulation. If the needs of the tissues for oxygen are more than the acute autoregulatory mechanism can supply, a long-term autoregulatory mechanism, requiring several weeks to develop, can cause still more increase in blood flow. This mechanism causes the actual number of blood vessels to increase. For instance, when a blood vessel has become partially occluded, such as a coronary vessel of the heart, new vessels grow into the area of poor blood supply, returning the blood supply toward normal after a few weeks. Also, if a person ascends to a high altitude and remains there for many weeks, he gradually develops an increased number of blood vessels in all his tissues, thereby supplying them with an increased quantity of oxygen. Conversely, when he comes back to a lower altitude, some of the vessels actually disappear after several more weeks, thus illustrating that this long-term vascular phenomenon helps to autoregulate blood flow in response to the tissues' needs.

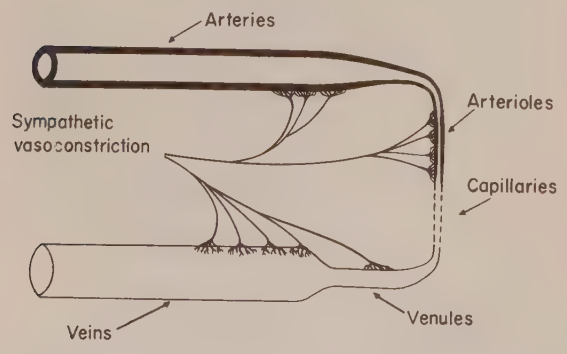


Figure 119. Innervation of the peripheral vascular system, showing especially the abundant innervation of the arterioles.

Nervous Control of Blood Flow

All the arterioles of the systemic circulation are supplied by nerves from the sympathetic nervous system, as illustrated in Figure 119. Stimulation of the sympathetic nerves causes the arterioles of most tissues to constrict.

In a few tissues, however, sympathetic nerves dilate the arterioles rather than constrict them, for instance in some areas of the skin and in the muscles as pointed out above. However, since by far the most important function of sympathetic control of blood vessels is vasoconstriction, the remainder of our discussion concerns the vasoconstrictor sympathetic fibers.

Vasoconstrictor tone of the systemic vessels. The sympathetic vasoconstrictor nerves normally transmit a continual stream of impulses to the blood vessels, maintaining the vessels in a moderate state of constriction all of the time. This is called *vasomotor tone*. The mechanism by which the sympathetic nervous system constricts vessels more than their normal state of constriction is to increase the number of sympathetic impulses above normal, while the mechanism by which it dilates the vessels is to decrease the impulses. In this manner the vasoconstrictor sympathetic system can cause both vasoconstriction and vasodilatation, which is the reason vasodilating nerves are not essential in most parts of the circulation.

Function of nervous control of the blood vessels. Nervous control of blood vessels is not generally concerned with regulating local nutrition of the tissues. Instead, it is concerned with the overall distribution of blood flow to major sections of the body. For instance, when one exercises, the muscles require such an extreme increase in blood flow that the heart often cannot pump an adequate quantity of blood both through the dilated muscle vessels and at the same time through such areas as the skin, the kidneys, and the gastrointestinal tract, which normally receive three fourths of the total cardiac output. Therefore, during exercise, sympathetic impulses cause vasoconstriction in these areas while the muscle vessels become dilated by the autoregulation and sympathetic vasodilator mechanisms. As a result, a major portion of the total blood flow shifts to the mus-

cles, thereby allowing much more muscular activity than could otherwise occur.

The sympathetic nervous system can also shift large amounts of blood flow to help regulate body temperature. When the body temperature rises too high, sympathetic control dilates the arterioles of the skin, increasing the flow of warm blood to the skin; this promotes heat loss until the temperature returns to normal. On the other hand, when the body temperature falls below normal, the opposite effects occur: blood flow decreases, heat loss becomes less, and body temperature rises.

A third instance of important nervous regulation of blood flow occurs when the circulatory system has been damaged so greatly that the heart pumps an insufficient quantity of blood to supply all portions of the body. Under these conditions sympathetic stimulation causes vasoconstriction in the less vital areas, while organs such as the brain and heart continue to receive adequate flow.

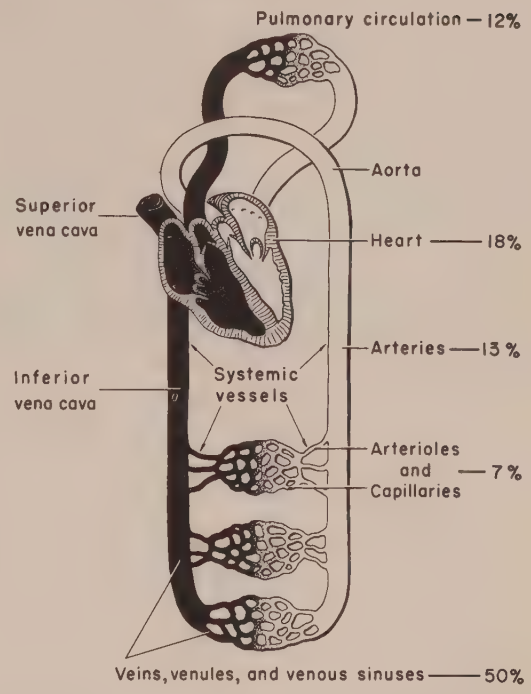


Figure 120. Distribution of blood in the different portions of the circulatory system.

DISTRIBUTION OF BLOOD IN THE BODY

Figure 120 shows the approximate percentages of the total blood in the different portions of the circulatory system. It is evident from this figure that about three fourths of the blood is in the systemic circulation and one-fourth in the pulmonary circulation and heart. Also, a far higher proportion of the blood is in the veins than the arteries. For instance, the arteries of the systemic circulation contain only 13 per cent of the total blood, while the veins, venules, and venous sinuses contain approximately 50 per cent. The arterioles and capillaries, despite their extreme importance to the circulation, contain only a few per cent of the blood.

The Blood Reservoirs

When the total blood volume falls so low that the vessels are no longer adequately filled, blood cannot circulate normally through the tissues. For this reason, it is important to have an extra supply of blood. The entire venous system acts as a *blood reservoir*, for the veins exhibit a plastic quality whereby their walls can distend and contract in response to the amount of blood available in the circulation. Also, the veins are supplied with nerves from the sympathetic nervous system so that any time the tissues begin to suffer for lack of blood flow, nervous reflexes cause a large number of sympathetic impulses to pass to the veins, constricting them and translocating blood into the heart and other vessels. It is this contractile and expansile quality of the venous system that protects the circulation against the disastrous effects of blood loss.

Certain portions of the venous system are especially important for storing blood:

First, the *large veins of the abdominal region* are particularly distensible, and, therefore, normally hold a tremendous amount of blood. Yet they can contract when the blood is needed elsewhere in the circulation.

Second, the *venous sinuses of the liver* can expand and contract many-fold, so that the liver under certain circumstances may hold as much as a liter and a half of blood, but at other times only a few hundred milliliters.

Third, the *spleen* normally contains approximately 200 ml. of blood, but can expand to hold as much as one liter, or can contract to hold as little as 50 to 100 ml.

The *venous plexuses of the skin* are a fourth very important blood reservoir. Normally the blood in these plexuses is used to regulate the heat of the body—the more rapidly blood flows through them the greater is the loss of heat. However, when the vital organs need extra blood flow, the sympathetic nervous system can markedly contract the skin's venous plexuses, transferring the stored blood into the main stream of flow.

A fifth blood reservoir is the *pulmonary vessels.* Approximately 12 per cent of the blood is normally in the pulmonary circulation, and much of this can be displaced into other portions of the circulation without impairing the function of the lungs. Therefore, the lungs act as a major source of blood in times of need.

REFERENCES

Alexander, R. S.: The systemic circulation. *Ann. Rev. Physiol.* 25:213, 1963.

Cooper, K. E.: The peripheral circulation. *Ann. Rev. Physiol.* 24:139, 1962.

Crismon, J. M.: Peripheral circulation. *Ann. Rev. Physiol.* 22:317, 1960.

Dobbing, J.: The blood-brain barrier. *Physiol. Rev.* 41:130, 1961.

Green, H. D.: Circulation: physical principles. In Glasser, O., editor: *Medical Physics.* Vol. I. Chicago, Year Book Publishers, Inc., 1944.

Green, H. D., and Kepchar, J. H.: Control of peripheral resistance in major systemic vascular beds. *Physiol. Rev.* 39:617, 1959.

Guyton, A. C.: Peripheral circulation. *Ann. Rev. Physiol.* 21:239, 1959.

Korner, P. I.: Circulatory adaptations in hypoxia. *Physiol. Rev.* 39:687, 1959.

SPECIAL AREAS OF THE CIRCULATORY SYSTEM

14

THE PULMONARY CIRCULATION

The pulmonary circulation is the circulation in the lungs. Its function is to transport blood through the pulmonary capillaries into which oxygen is absorbed from the alveolar air and from which carbon dioxide is excreted.

The physiologic anatomy of the pulmonary circulation, illustrated in Figure 121, is very simple. The right ventricle pumps blood into the pulmonary artery. From there, the blood flows through the pulmonary capillaries into the pulmonary veins and finally into the left atrium. Because all portions of the lungs have the same function—that is, to aerate the blood—the arrangement of the vessels is the same in all areas of the pulmonary circulation.

The pulmonary capillaries lie immediately adjacent to the epithelial lining of the alveoli. The combination of the alveolar lining, capillary membrane, and the small amount of tissue that lies between these two is called the *pulmonary membrane,* the total thickness of which is only 0.2 to 0.4 micron. The pores of this membrane are large enough for oxygen and carbon dioxide to diffuse through them with ease. Yet they are small enough that the blood proteins do not leak out of the capillaries.

Flow of Blood through the Lungs

Because blood flows around a continuous circuit in the body, the same amount of blood

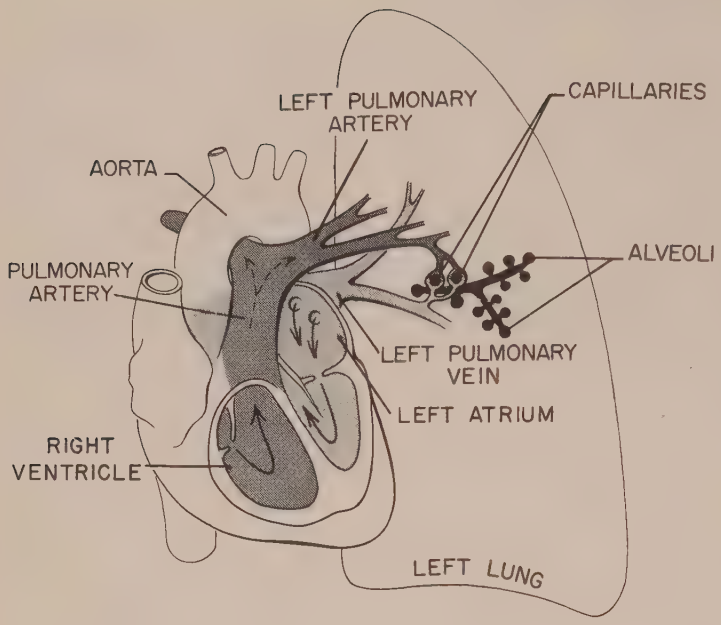


Figure 121. The pulmonary circulation and its relation to the heart.

must flow through the lungs as through the systemic circulation. The vessels of the lungs are very expansile so that whenever the amount of blood entering the lungs increases, the pulmonary vessels passively dilate to allow the more rapid flow demanded of the pulmonary circulation. In other words, the resistance to blood flow through the lungs decreases as the rate of blood flow increases. This allows facile transport of blood through the lungs under widely varying conditions.

Regulation of blood flow through the lungs. Since all portions of the lungs perform essentially the same function, there is no need for extensive regulation of the pulmonary blood vessels. Instead, the flow through the pulmonary circulation is great when it is great in the systemic circulation and slight when it is slight also in the systemic circulation.

However, one feature of blood flow regulation in the lungs is different from that in the remainder of the body. It is quite important for blood to flow only through those portions of the lungs that are adequately aerated and not through portions not aerated. Research studies in the past few years have shown that a low level of oxygen in one area of the lungs automatically causes the vessels there to constrict. Therefore, when the bronchi to an area of the lungs become blocked, the alveolar oxygen is rapidly used up, and as a consequence, the vessels constrict, thereby forcing the blood to flow through areas of the lungs that are still aerated.

Pressures in the Pulmonary Circulation

The resistance to blood flow in the pulmonary circulation is so slight that the mean pulmonary arterial pressure is only 13 mm. Hg and the systolic pressure (the highest pressure during contraction of the ventricle) is only 22 mm. Hg, while the diastolic is 8 mm. Hg (the lowest pressure between contractions). The mean pulmonary arterial pressure is approximately one-seventh the mean systemic arterial pressure which is about 100 mm. Hg, as is discussed in the following chapter.

The pulmonary venous pressure is about 4 mm. Hg. This is also the pressure in the left atrium. The pulmonary capillary pressure has never been measured, but it must be greater than the pulmonary venous pressure and less than the pulmonary arterial diastolic pressure—that is, somewhere between 4 and 8 mm. Hg. Indirect measurements of pulmonary capillary pressure indicate that it is probably about 5 mm. Hg.

The mean pressure gradient from the pulmonary artery to the pulmonary vein is 13 mm. Hg minus 4 mm. Hg or 9 mm. Hg. This compares with a pressure gradient in the systemic circulation of approximately 100 mm. Hg. To express this differently, the same amount of blood flows through the lungs as through the systemic circulation with about one-eleventh the force propelling it. This means also that the total resistance of the pulmonary circulation is normally only one-

eleventh the total resistance of the systemic circulation.

Pressure in the pulmonary artery during exercise. When one exercises, the amount of blood flowing through the systemic circulation sometimes increases to as high as five times normal, which means that the amount flowing through the lungs must also increase this much. Still, the pulmonary arterial pressure does not increase more than 25 to 30 per cent because the vessels of the lungs passively enlarge to accommodate whatever amount of blood needs to flow, which keeps the pressure from rising more than a few millimeters of mercury. This obviously keeps the right heart from becoming overloaded during exercise.

Pulmonary capillary dynamics. Only about one tenth of the blood of the lungs is actually in the capillaries at any one time, and the blood flows through them so rapidly that it remains in the capillaries at most only one to two seconds. During exercise, when blood flow is very rapid, the blood may remain in the pulmonary capillaries only one-quarter to one-half second. Yet, because the pulmonary membrane is extremely porous, even during this very minute period of time the blood can become almost completely oxygenated and can excrete the necessary amount of carbon dioxide.

CAPILLARY MECHANISM FOR MAINTAINING DRY LUNGS. Another extremely important feature of pulmonary capillary dynamics is the mechanism for keeping the air spaces of the lungs dry. The plasma of the blood has a colloid osmotic pressure of 28 mm. Hg, which causes a continual tendency for fluid to be absorbed into the capillaries. The mean pulmonary capillary pressure is only 5 mm. Hg. Therefore, the tendency for fluid to be absorbed into the capillaries is about 23 mm. Hg greater than for fluid to filter out of the capillaries. Thus, the plasma colloid osmotic pressure continually promotes fluid absorption from the tissues and alveoli of the lungs, and if a small amount of water does enter the alveoli, it is usually absorbed within a few minutes. Indeed, when a person drowns, the absorption of water is so rapid that more than a liter is absorbed in the 2 to 3 minutes before he dies.

Abnormalities of the Pulmonary Circulation

Pulmonary congestion means too much blood and fluid in the lungs, and it results from too high a pressure in the pulmonary circulation, which in turn is caused most frequently by failure of the left heart to pump blood adequately from the lungs into the systemic circulation. This causes blood to become dammed up in all of the pulmonary vessels, greatly increasing their pressures. Once the pulmonary capillary pressure rises above the plasma colloid osmotic pressure, about 28 mm. Hg, fluid leaks out of the capillaries extremely rapidly, and the alveoli no longer remain dry. Thus, severe *pulmonary edema* develops, sometimes so rapidly that it causes death in 1 to 2 hours.

Atelectasis means collapse of a lung or part of a lung. Occasionally some abnormality blocks one or more of the bronchi, cutting off air flow. When this happens, the air in the blocked alveoli is absorbed within a few hours, causing the lung to collapse. Simultaneously, the lack of oxygen in the collapsed alveoli causes vascular constriction, a mechanism that was discussed earlier in the chapter. As a result, the blood flow through the collapsed lung decreases to about one-fifth its previous value, which causes most of the blood to pass through the uncollapsed areas of the lungs where it can be aerated.

Surgical removal of large portions of the lungs causes excessive blood flow through the remaining lung tissue. Ordinarily blood can flow through each lung four times as rapidly as normally before the pulmonary arterial pressure begins to rise. Therefore, a person can lose one whole lung, which increases the blood flow through the opposite lung about two-fold, without causing any serious disability. Yet if this person attempts too heavy exercise, he rapidly reaches the upper limit that the blood can flow before pulmonary hypertension and failure of the right heart develop.

Effect of Congenital Heart Disease on the Pulmonary Circulation

Congenital heart disease means an abnormality of the heart or of closely allied blood vessels that is present at birth. Many types of congenital heart disease affect the pulmonary circulation in particular; two of these are illustrated in Figures 122 and 123.

During normal fetal life, blood by-passes the lungs by flowing from the pulmonary artery through the ductus arteriosus into the aorta. Immediately after birth, expansion of the lungs dilates the pulmonary vessels so that blood can then flow through them with ease. As a result, pulmonary arterial pressure falls, causing the blood to flow backward from the aorta into the pulmonary artery. This backward flow causes the ductus arteriosus to become occluded during the first few days or weeks of life in almost all babies. In perhaps 1 out of every 2000 the ductus never closes—that is, the *patent ductus arteriosus,* an open ves-

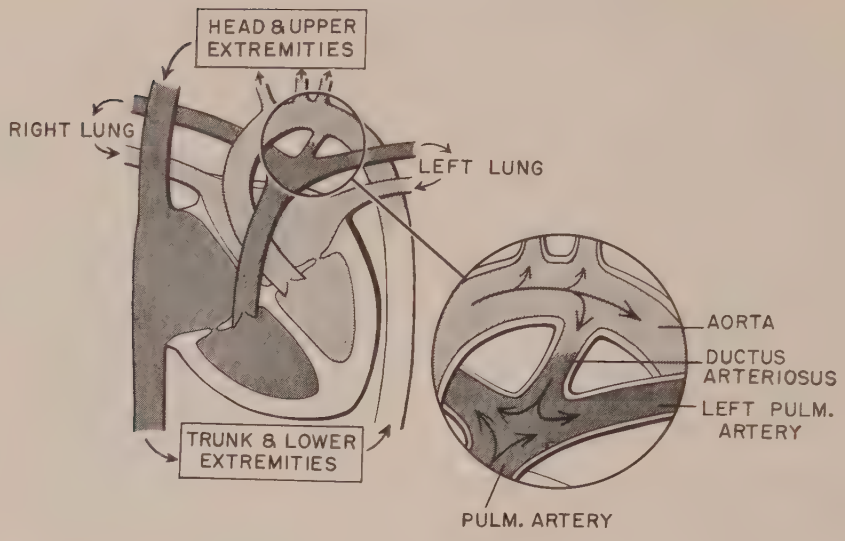


Figure 122. Patent ductus arteriosus.

sel connecting the aorta and the pulmonary artery, is present throughout life (see Figure 122). As the child grows older and his heart becomes stronger, the pressure in the aorta becomes much greater than the pressure in the pulmonary artery, forcing a tremendous quantity of blood from the aorta into the pulmonary artery. This blood flows a second time through the lungs, then back through the left heart, through the lungs again, and around and around this circuit several times before finally passing into the systemic circulation. In other words, many times the normal amount of blood is pumped through the lungs. In early life this is a great strain on the heart but does not cause extensive damage to the lungs. However, in early adulthood the very high pressure in the pulmonary vessels eventually promotes fibrosis of the vessels and congestion of the lungs. Finally, the person has such difficulty aerating his blood that this alone may cause his death.

Figure 123 shows the congenital abnormality called *tetralogy of Fallot.* This disease presents four abnormalities of the heart, including (1) a hole in the septum (partition) between the right and left ventricles, (2) a greatly constricted pulmonary artery, (3) displacement of the aorta to the right, and (4) a hypertrophied right ventricular muscle. Very little blood passes from the right ventricle through the constricted pulmonary artery. Instead, most of it is forced through the

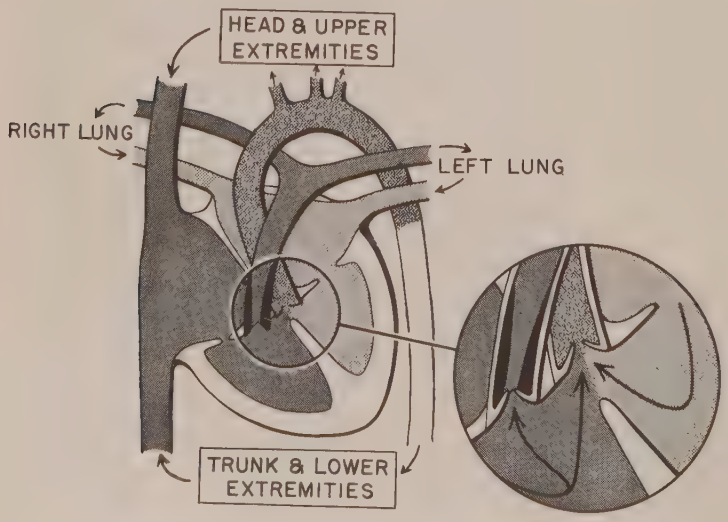


Figure 123. Tetralogy of Fallot.

hole in the septum and then into the aorta. Thus, most of the blood pumped by the right ventricle by-passes the lungs and is not aerated. Because non-aerated blood is dark bluish-red, the whole body assumes this hue. Tetralogy of Fallot is one of the congenital anomalies of the heart that causes the so-called *blue baby.*

Other often encountered congenital abnormalities of the heart are: (1) an abnormal opening through the ventricular septum, (2) an abnormal opening through the atrial septum, (3) constriction or occlusion of the descending aorta, so that blood must flow to the lower part of the body through many small vessels in the body wall rather than through the aorta, (4) abnormal connection of one of the pulmonary veins to the right atrium instead of to the left atrium, and (5) partial constriction of the pulmonary artery, resulting in high right ventricular pressure and overloading of the right ventricle.

In the past few years many advances have been made in the surgical repair of cardiac abnormalities. For instance, any patient with a patent ductus arteriosus can be treated very simply by placing a tight ligature around the ductus. Treatment of tetralogy of Fallot is more complicated but is usually accomplished reasonably satisfactorily by creating an artificial patent ductus; an opening is made between one of the major systemic arteries and the pulmonary artery, allowing much of the blood that normally by-passes the lungs to flow into the pulmonary system. Surgical procedures have also been devised to transpose blood vessels when necessary and to close abnormal openings in the ventricular or atrial septum.

THE CORONARY CIRCULATION

Contrary to what might be expected, blood does not pass directly from the chambers of the heart into the heart muscle. Instead, the heart has its own special circulatory system, which is shown in Figure 124. Two small arteries called the *coronary arteries* originate from the aorta immediately beyond the aortic valve. Blood flows from these vessels through branches over the *outer surface* of the heart, then into smaller arteries and capillaries in the cardiac muscle, and finally to the right atrium mainly through a very large vein called the *coronary sinus.*

Flow of Blood through the Coronary Vessels

The amount of blood flowing through the coronaries each minute is approximately 225 ml., which amounts to approximately 4 to 5 per cent of all the blood pumped by the heart.

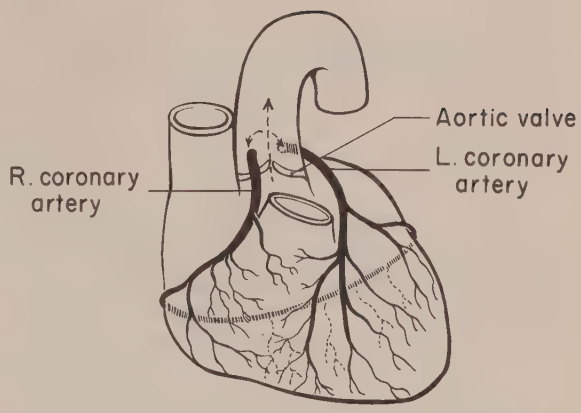


Figure 124. The coronary circulation.

In contrast to the flow of blood in other portions of the circulation, flow in the coronaries is greater during diastole than during systole. The reason for this is that systolic contraction of the heart muscle compresses the vessels in the wall of the heart. This momentarily partially or totally occludes these vessels. Therefore, the rate of coronary flow is determined much more by the diastolic level of arterial pressure than by either the mean pressure or the systolic pressure. This has importance in those circulatory diseases in which diastolic pressure is especially low.

Regulation of coronary blood flow. Blood flow through the coronaries is regulated in proportion to the need of the heart for nutrition. The most important method of regulation is the *autoregulation mechanism,* which was discussed in the previous chapter. That is, when the metabolic need for nutrients—especially for oxygen—becomes greater than the supply, the arterioles automatically dilate. As a result, blood flow through the coronaries increases until the level of nutrition matches the degree of cardiac activity.

The sympathetic nervous system also helps slightly to control blood flow through the coronaries. Increased sympathetic stimulation increases coronary flow and decreased sympathetic stimulation decreases flow. Physiologists are not sure, however, whether this is a direct effect of sympathetic stimulation or an indirect effect as follows: Sympathetic stimulation increases the degree of activity of the cardiac muscle tremendously which in turn would be expected to increase coronary blood flow by the autoregulation mechanism discussed previously. Therefore, some physiologists believe that the increased coronary flow following sympathetic stimulation is simply another manifestation of autoregulation.

However, nervous control of coronary blood flow is relatively unimportant in comparison with the autoregulatory control of flow.

Coronary Occlusion

Coronary occlusion means blockage of one of the coronary vessels. When this occurs that portion of the heart formerly supplied by the vessel loses its nutrition, stops contracting, and becomes so flaccid that it actually dilates outward during the period of the cardiac cycle when the pressure inside the heart becomes great. Obviously, when a major portion of the heart is affected, its pumping capability becomes so depressed that it can no longer pump sufficient blood to maintain life.

Coronary occlusion is the cause of approximately one third of all deaths. Occlusion may occur rapidly or slowly, and in almost all instances the cause is *atherosclerosis*. Therefore, to explain why people have heart attacks, it is first necessary to discuss the cause and the effects of atherosclerosis.

Atherosclerosis. Atherosclerosis is a disease of fat metabolism in which the body fails to utilize fat in the normal manner. As a result, fatty deposits that contain mainly cholesterol, plus smaller amounts of phospholipids and neutral fat, appear in the walls of the arteries. Gradually, fibrous tissue grows around or into the fatty deposits, and calcium from the body fluids frequently combines with the fat to form solid calcium compounds, eventually evolving into hard plates similar to bone. Thus, in the early stage of atherosclerosis, only fatty deposits occur in the walls of the vessels, but in the late stage the vessels may become extremely fibrotic and constricted, or even bony hard in consistency, a condition called "hardening of the arteries."

Acute coronary occlusion. Atherosclerosis can cause acute occlusion of the coronary vessels in three ways: First, the fatty deposits may break through the inside surface of the vessel and cause clotting of the blood, which in turn plugs the vessel. Second, the protruding fat may actually break away from its original deposit and flow into and occlude a smaller vessel. Third, the fatty deposits may erode into one of the tiny blood vessels called *vasa vasorum*, which supply the walls of the coronaries. When this occurs the minute vessel bleeds into the wall of the coronary pushing the inner lining inward, thereby totally or partially blocking blood flow. In any of these instances, the resulting coronary occlusion can cause sudden diminishment of cardiac function, which is the familiar condition known as a "heart attack." If the occlusion is extremely severe, it will cause immediate death; if it affects only a small vessel, the heart is often only temporarily weakened for one to three months until some of the vascular connections with neighboring blood vessels enlarge to supply new blood. After a heart attack, the heart sometimes returns to full strength, though more often it remains weakened for life.

Slowly developing occlusion. The slowly developing type of coronary occlusion results from fibrous tissue invading the atherosclerotic fatty deposits. The fibrous tissue contracts slowly and continually, gradually narrowing the vessel. This effect occurs in all people as they become old, though it occurs in some at much earlier ages than in others. Actually, coronary occlusion eventually occurs in everyone who lives long enough. Indeed, almost 100 per cent of hearts examined after death of persons beyond the age of 60, whether the persons had ever had obvious heart attacks or not, show at least one fairly large coronary vessel totally occluded. Ordinarily, accessory blood vessels will have grown into the area supplied by the occluded vessel so that the person may never have known of his coronary disease. The point of this discussion is that all people develop coronary heart disease, which means that it is important for each person as he becomes old to treat himself as if he had heart disease.

Relationship of diet to coronary disease. Recent studies have shown that atherosclerosis, and consequently heart attacks as well, can be minimized by two dietary measures; First, the total intake of food should be little enough for the weight to remain either normal or even slightly subnormal. Second, the amount of fat in the diet should be as little as possible. The average American and northwestern European diets contain far more fat than those of other sections of the world, and as a result the incidence of heart attacks is far higher than among such people as the Italians, Japanese, and Chinese. Statistics have also shown that smoking almost doubles the incidence of heart disease, but why this occurs is not known.

THE CEREBRAL CIRCULATION

Figure 125 illustrates the arterial system of the brain, showing the two *vertebral arteries* passing upward along the medulla and the two *internal carotid arteries* approaching the brain on its ventral surface. Communicating branches between these four major arteries protect the brain against damage should any one or even two of them become occluded. However, the intermediate-size arteries which spread over the surface of the

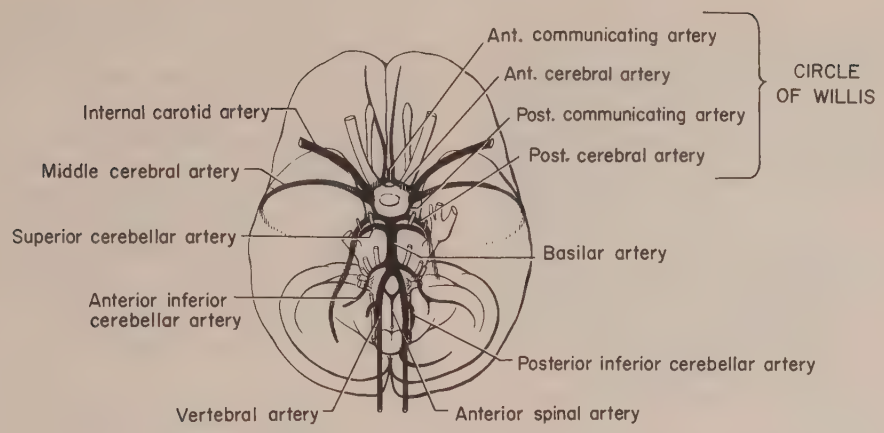


Figure 125. The cerebral circulation.

brain do not have many communications from one to another. Therefore, occlusion of any one of them usually causes destruction of the respective brain tissue.

Cerebral Blood Flow and Its Regulation

The total amount of blood flowing to the brain averages approximately 650 to 700 ml. per minute, which is about 13 per cent of the blood pumped by the heart.

Autoregulation of cerebral blood flow. Cerebral blood flow is autoregulated better than that in almost any other area of the body. Even though the arterial pressure might fall to as low as 40 mm. Hg or rise to as high as 200 mm. Hg, the blood flow through the brain still varies no more than a few per cent.

Autoregulation of blood flow in the brain results from two different mechanisms: (1) the oxygen deficiency mechanism, which was described in the previous chapter and (2) a carbon dioxide autoregulation mechanism. The oxygen deficiency mechanism operates in the same manner as in all other tissues of the body. That is, lack of oxygen causes vasodilation, whereas excess of oxygen causes vasoconstriction. However, the cerebral circulation is different from that of other tissues in that it responds to carbon dioxide concentration as well as to oxygen concentration. When the carbon dioxide concentration rises above normal, cerebral blood flow automatically increases and washes carbon dioxide out of the tissues, bringing the carbon dioxide concentration back toward normal. Conversely, decreased carbon dioxide in the brain decreases the blood flow, which again returns the carbon dioxide concentration toward normal.

The carbon dioxide mechanism is much more powerful in the cerebral circulation than the oxygen mechanism, and it is mainly because of this mechanism that cerebral blood flow is autoregulated to an almost exactly constant level all the time. The only exception to this occurs when the brain itself becomes overly active. Under these conditions cerebral blood flow can on rare occasions rise to as high as 30 to 50 per cent above normal, but even this increase in flow is a manifestation of autoregulation, the flow increasing in proportion to the increase in activity of the brain.

The unvarying rate of cerebral blood flow is very advantageous to cerebral function, for neurons must never become excessively excitable nor excessively inhibited if the brain is to function properly, and, obviously, changes in the nutritional status of the neurons could quite easily change their degree of excitability. For instance, if the carbon dioxide concentration in the tissues should rise too high, the tissue fluids would become excessively acidic, which is known to depress neuronal function.

Venous Dynamics in the Cerebral Circulation

Function of the veins in the cerebral circulation is somewhat different from that of veins in the remainder of the body because the cerebral veins are non-collapsible. They are formed as *sinuses* in the fibrous lining, the *dura mater,* of the cerebral vault, and the sides of the sinuses are so firmly attached that they cannot collapse. Also, because the head is in an upright position most of the time, the weight of the venous blood makes it run rapidly downward toward the heart, creating a slightly negative pressure in the cerebral veins. If one of these veins is inadvertently punctured,

air is often sucked into the venous system and occasionally may even flow so rapidly to the heart that it blocks the action of the heart valves and kills the person. Because the veins in all other portions of the body are flimsy enough to collapse, any sucking action that might occur simply closes the veins rather than creating a negative pressure.

Cerebral Vascular Accidents—the "Stroke"

When a blood vessel supplying blood to an important part of the brain suddenly becomes blocked or ruptures, the person is said to have had a cerebral vascular accident or a "stroke." The incidence of this is high. In about one fourth of all cases the damage is caused by a blood clot developing on an atherosclerotic fatty deposit in a major cerebral artery. In the other three-fourths, an artery ruptures because of excessively high arterial pressure or because the vessel itself has been weakened by some disease process such as atherosclerosis. Such hemorrhages into the tissues of the brain often compress the neuronal cells enough to destroy them or at least to stop their functioning. Also, the blood clot developing in the hemorrhagic area extends backward into the artery, blocking further flow along the inside of the vessel. Therefore, regardless of whether the cerebral vascular accident is initially caused by simple occlusion or by hemorrhage, the results are essentially the same.

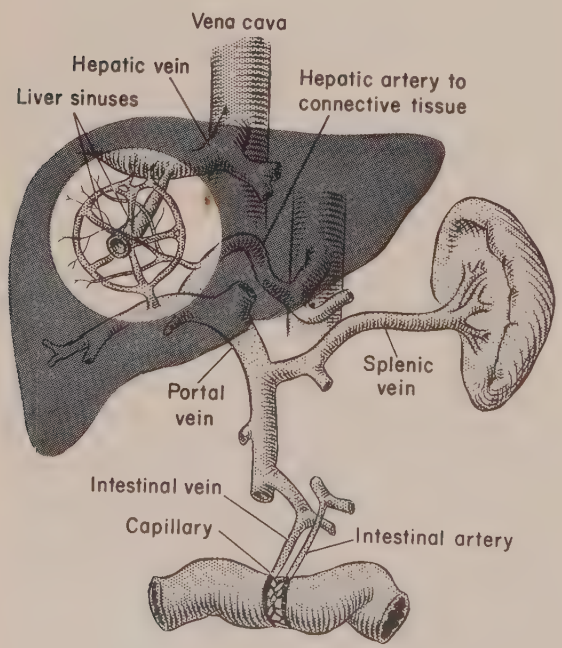


Figure 126. The portal circulation.

The effects of a cerebral vascular accident depend on which vessel is destroyed. Frequently it is the middle cerebral artery supplying the area of the brain that controls muscular function, in which case the opposite side of the body becomes paralyzed. Blockage of the posterior cerebral artery, on the other hand, causes partial blindness. Blockage of arteries in the hindbrain region is likely to destroy nerve tracts connecting the brain and spinal cord, causing any number of abnormalities such as paralysis, loss of sensation, and loss of equilibrium.

THE PORTAL CIRCULATORY SYSTEM

The portal circulatory system, shown in Figure 126, is composed of the veins leading from the intestines and spleen to the liver and thence through the liver into the vena cava. The function of this system is to pass the blood from the intestines through the liver prior to its entering the general circulation. On flowing through the minute sinuses of the liver the blood comes into contact with special phagocytic cells, the *Kupffer cells,* that are capable of removing abnormal debris. These cells, therefore, cleanse the intestinal blood before it reenters the general circulation. This is quite important because each minute a few of the literally billions of bacteria in the gastrointestinal tract make their way into the portal blood. The Kupffer cells are so effective in removing the bacteria that probably not one in a thousand escapes through the liver into the general circulation.

The portal system also allows the liver to remove various absorbed nutrients from the intestinal blood before they reach the general circulation. For instance, the liver cells normally remove approximately two thirds of the glucose absorbed into the portal blood from the intestines and perhaps as much as one half of the proteins before the blood ever reaches the general circulation. These nutrients are stored in the liver until they are needed later by the remainder of the body and then are released back into the general circulation. The storage of glucose and proteins "buffers" the blood concentrations of these nutrients, for it prevents sudden surges in their concentrations in the blood and also prevents low concentrations during the long intervals between meals.

Vascular Dynamics of the Portal System

The small vessels in the liver offer so much resistance to blood flow that the portal venous

pressure is normally about 7 mm. Hg greater than the pressure in the vena cava. Ordinarily this elevated portal pressure is of no major concern, but when some abnormality causes the systemic venous pressure to rise, the portal pressure increases a corresponding amount, always remaining about 7 mm. Hg greater than the systemic venous pressure. As a result, the portal capillary bed is one of the first to suffer from excessively high systemic venous pressures, a condition that occurs most frequently when a failing heart causes blood to dam up in the veins. The high capillary pressure forces fluid to leak through the capillary walls into the intestinal tissues and peritoneal cavity. Therefore, instead of the norm of only a few milliliters of fluid in the abdomen, many liters may accumulate. This is called *ascites*.

Portal obstruction. Occasionally the portal vein leading to the liver becomes totally occluded by a large blood clot, or more frequently a large proportion of the small vessels in the liver may become blocked because of liver disease. In either event, the portal venous pressure rises, which in turn causes high portal capillary pressure, edema of the intestinal walls, enlargement of the spleen, and ascites.

If the portal vein becomes occluded gradually, the normally minute venous channels that directly connect the portal and systemic veins, without going through the liver, become greatly enlarged. Veins passing from the stomach up along the esophagus sometimes enlarge so much that they erode through the inner lining of the esophagus and rupture, allowing the person to bleed severely, often to death. Enlarged veins of the anal region may develop into massive hemorrhoids which, too, can bleed severely.

If the portal vein becomes occluded suddenly, death usually ensues within a few hours, for then the accessory venous channels do not have time to develop. Instead, blood is dammed up in the portal venous system much more rapidly than it can be replenished. Extreme amounts of ascitic fluid containing large quantities of plasma protein develop. This, too, depletes the volume of blood in the general circulation, and the person dies of shock.

Function of the Spleen

Reservoir function. As was pointed out in Chapter 13, the spleen, which is part of the portal system, is a blood reservoir. It is capable of enlarging up to a total volume of more than 1000 ml. or of contracting down to a minimal volume of less than 100 ml. Its internal structure is illustrated in Figure 127, which shows several small arteries leading into capillaries in the *pulp* of the spleen and finally connecting with large venous sinuses. The blood flows from the venous sinuses into the large circumferential veins and then back into the general circulation. Much of the blood, including the red blood cells, leaks from the capillaries into the pulp and then diffuses through this area, finally reentering the general circulation through the walls of the venous sinuses. Large numbers of cells are often held in the pulp while the plasma is returned to the circulation. In this way the spleen stores a much higher percentage of red blood cells than plasma, but when it contracts, the cells are forced from the pulp back into the circulation. Therefore, the spleen is often said to be a red blood cell reservoir rather than a general blood reservoir.

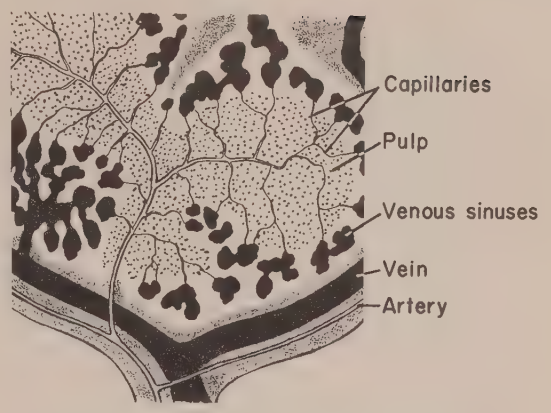


Figure 127. Internal organization of the spleen. (Modified from Bloom and Fawcett: *A Textbook of Histology*. 8th Ed.)

In lower animals, sympathetic stimulation causes smooth muscle in the capsule of the spleen to contract, which expels blood both from the splenic pulp and splenic sinuses into the general circulation. In the human being there is very little smooth muscle in the capsule. Therefore, it is believed that sympathetic stimulation causes the spleen of the human being to empty its blood mainly by direct constrictive effects on the vascular walls and perhaps other smooth muscle in the structure of the spleen itself.

Marked splenic contraction occurs during muscular exercise. The extra red blood cells expressed from the spleen into the circulation aid in the transport of oxygen to the active muscles.

Phagocytic function of the spleen. Another major function of the spleen is to cleanse the blood. This results from phagocytosis by reticuloendothelial cells that line the splenic sinuses or are located throughout the splenic pulp.

MUSCULAR BLOOD FLOW

Under resting conditions the blood flow through the muscles is only 15 to 20 per cent of the total flow in the body. During exercise, muscular blood flow increases markedly in response to the increase in muscular metabolism—that is, in response to the increase in use of nutrients by the muscle cells. During extreme exercise, muscular blood flow can increase 15-fold, and sometimes even more than this.

Mechanism of Blood Flow Regulation in the Muscles

Autoregulation. The principal mechanism by which blood flow is regulated in the muscles is the autoregulation mechanism which was described in the previous chapter. Increased muscle metabolism increases the utilization of nutrients from the blood, including oxygen, which in turn has a direct vasodilator effect on the blood vessels to increase the rate of blood flow. It is possible that the diminished oxygen causes vasodilator substances to be formed in the muscle, but another likely effect is simply that lack of oxygen directly causes the blood vessels to dilate.

Sympathetic vasodilator nerves to the muscles. A second mechanism that dilates the muscle vessels during exercise is stimulation of sympathetic vasodilator nerves. These are controlled by areas of the cerebral cortex in the brain near those that control contraction of the exercising muscles. When the muscles are contracted, impulses are sent simultaneously to the sympathetic vasodilator nerves, thereby increasing the blood flow. However, the importance of the vasodilator fibers has yet to be proved because removal of all sympathetic nerves to muscles hardly affects the degree of vasodilation that occurs during exercise. It is believed, though, that the sympathetic vasodilator nerves perform the following function: The autoregulation mechanism requires 10 to 30 seconds to develop fully after exercise begins. On the other hand, the sympathetic vasodilator mechanism can dilate the blood vessels within the first second or more after exercise begins, and sometimes even *before* exercise begins. Therefore, it is believed that the sympathetic vasodilator mechanism provides an immediate means for increasing blood flow at the onset of muscular exercise, thereby anticipating the increased needs of the muscles for nutrients rather than waiting for a deficiency to occur, as is the case with the autoregulation mechanism.

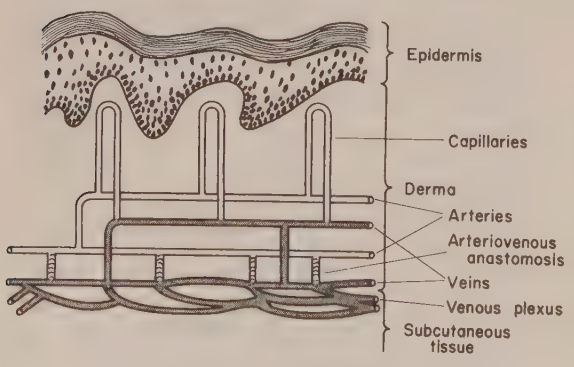


Figure 128. The circulation in the skin.

Effect of increased arterial pressure. Another factor that increases muscle blood flow during exercise is increased arterial pressure. During very heavy exercise the mean arterial pressure rises from about 100 mm. Hg to 150 or occasionally even to 175 mm. Hg. This increased pressure obviously forces increased amounts of blood through the muscle blood vessels. Also, sympathetic stimulation, which is very powerful during muscle exercise, constricts essentially all the other vessels of the body, thereby shunting the blood flow out of other areas and through the muscles.

BLOOD FLOW THROUGH THE SKIN

Blood flow through the skin has two functions: first, to regulate the temperature of the body and, second, to supply nutrition to the skin itself.

Relationship of skin blood flow to body temperature regulation. Figure 128 shows an extensive venous plexus lying a few millimeters beneath the surface of the skin. The rate of blood flow through this plexus can probably be altered as much as 100-fold. When the arteries that supply blood to the venous plexus are constricted, the blood flow may be as little as 20 to 50 ml. per minute to the plexus of the entire body, whereas, when the arteries are completely dilated, the blood flow through the plexus may be as great as 2 to 3 liters per minute.

Obviously, if the blood flow to the venous plexus is very rapid, the amount of heat transmitted by the blood from the internal structures to the surface of the body will be great, and large amounts of heat will be radiated from the skin. Conversely, when the quantity of blood flowing to the plexus is slight, the amount of heat lost from the body will also be slight. Special temperature regulatory mechanisms, controlled by the hypothalamus in the brain and acting through the sympathetic nerves, can either vasoconstrict or vasodilate the skin vessels, thereby helping to

regulate body temperature. This will be discussed in much more detail in Chapter 32.

Certain areas of the body, such as the hands, feet, and ears have special structures, the *arteriovenous anastomoses,* which connect the arteries directly with the venous plexus. When these anastomoses are dilated, blood bypasses the capillary system and flushes extremely rapidly into the plexus. In this way the hands, feet, and ears receive tremendous amounts of blood flow when they are exposed to excessive cold, the increased blood flow obviously acting to protect the tissues from freezing.

The nutrient blood flow. Ordinarily the blood flow through the skin is some 20 to 30 times that required to supply sufficient nutrition to the skin tissues. Yet, when the skin becomes extremely cold the body temperature mechanisms constrict the skin blood flow so greatly that nutrition can become impaired. It is only under such conditions as this that one need be concerned with the nutrient vascular supply.

Figure 128 illustrates capillary loops that transmit blood into all the tufts beneath the skin. The nutritive blood flow is subject to the same local autoregulatory mechanism for blood flow as that found elsewhere in the body. For instance, when a person compresses an area of the skin so hard that the blood flow is completely blocked for a few minutes, the skin flow becomes far greater than normal immediately after relief of the pressure, which automatically makes up for the deficiency in nutrients. Fortunately, the skin can be without blood flow for as long as an hour or more before serious damage takes place.

REFERENCES

Cournand, A.: Some aspects of the pulmonary circulation in normal man and in chronic cardiopulmonary diseases. *Circulation* 2:641, 1950.

Friedberg, C. K.: *Diseases of the Heart.* 2nd Ed. Philadelphia, W. B. Saunders Company, 1956.

Gofman, J. W.: *Coronary Heart Disease.* Springfield, Ill., Charles C Thomas, 1959.

Gregg, D. E.: *Coronary Circulation in Health and Disease.* Philadelphia, Lea & Febiger, 1950.

Hertzman, A. B.: Vasomotor regulation of cutaneous circulation. *Physiol. Rev.* 39:280, 1959.

Lassen, N. A.: Cerebral blood flow and oxygen consumption in man. *Physiol. Rev.* 39:183, 1959.

Lindsey, A. W., and Guyton, A. C.: Continuous recording of pulmonary blood volume, and pulmonary pressure and volume changes during acute right or left ventricular failure. *Am. J. Physiol.* 197:959, 1959.

Scherf, D., and Boyd, L. J.: *Cardiovascular Diseases.* 3rd Ed. New York, Grune & Stratton, Inc., 1958.

Schmidt, C. F.: *The Cerebral Circulation in Health and Disease.* Springfield, Ill., Charles C Thomas, 1950.

Wright, I. S., and Millikan, C. H.: *Cerebral Vascular Diseases.* New York, Grune & Stratton, Inc., 1958.

SYSTEMIC ARTERIAL PRESSURE AND HYPERTENSION

15

When the left ventricle contracts it forces blood into the systemic arteries, thereby creating pressure that drives the blood through the systemic circulation. The *regulation of arterial pressure* is among the most important subjects of circulatory physiology, and the disease *hypertension*, which means high arterial pressure, is one of the most common of all disorders of the body. These two topics receive special emphasis in the present chapter.

PULSATILE ARTERIAL PRESSURE

Instead of pumping a continuous stream, the heart pumps a small quantity of blood with each beat. As a result, the arterial pressure rises during systole but falls during diastole. Figure 129 shows graphically this pulsatile pressure under both normal and abnormal conditions.

Systolic and Diastolic Pressures

The pressure at its highest point during the pressure cycle is called the *systolic pressure*, and the pressure at its lowest point is called the *diastolic pressure*. As shown in Figure 129, systolic pressure in a normal young adult is approximately

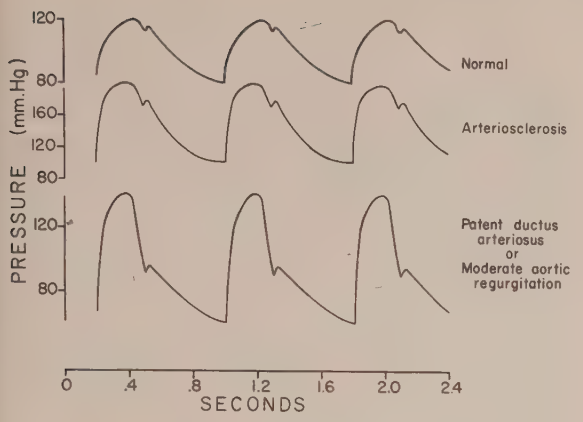


Figure 129. Pressure pulse contours in the normal circulation, in arteriosclerosis, in patent ductus arteriosus, and in moderate aortic regurgitation.

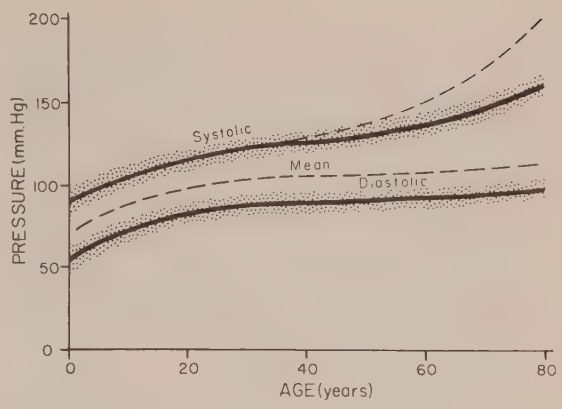


Figure 130. Changes in systolic, diastolic, and mean pressures with age. The shaded areas show the normal range. The upper dashed curve shows the effect of arteriosclerosis on systolic pressure in old age.

120 mm. Hg, while the diastolic pressure is approximately 80 mm. Hg. The usual method for writing these pressures is 120/80. As another example, if a person's blood pressure is stated to be 210/125 this means that the systolic pressure is 210 mm. Hg and the diastolic pressure 125 mm. Hg.

The systolic and diastolic pressures change with age, as demonstrated graphically in Figure 130. In a newborn baby the systolic pressure is about 90 mm. Hg and the diastolic about 55 mm. Hg. The pressure usually reaches 120/80 by young adulthood, and in old age an average of about 150/90. The upper dashed line of Figure 130 illustrates what happens to the systolic pressures when a person develops severe arteriosclerosis; this effect is explained in a later section of this chapter.

Measurement of systolic and diastolic pressures. The arterial pressure varies so rapidly during the cardiac cycle that special high-speed recorders designed to follow the rapid changes in pressure must be used to measure it. Several electronic transducers, used with appropriate electronic recorders, were described in Chapter 13 and illustrated in Figure 112. The basic principle of all these transducers is the following: The vessel from which the pressure is to be measured is connected by way of a non-distensible tube with a small chamber bounded on one side by a thin elastic membrane. When the pressure rises, the membrane bulges outward, and when the pressure falls the bulge decreases. The bulge in turn activates a variable capacitor, a variable resistor, or a variable inductor, the electrical output from which is amplified and recorded on a moving strip of paper.

Appropriately designed membrane transducers can record pressure changes that occur in less than one-hundredth of a second, which is more than adequate for recording any significant pressure change in the circulatory system. The arterial pressure pulse curves illustrated in Figure 129 were recorded using a membrane type electronic transducer.

INDIRECT MEASUREMENT OF SYSTOLIC AND DIASTOLIC PRESSURES BY THE AUSCULTATORY METHOD. Figure 131 illustrates the method usually employed in the doctor's office for measuring systolic and diastolic pressures. An inflatable blood pressure cuff is placed around the upper arm and is connected by a tube to a mercury manometer. When the pressure in the cuff is elevated above that in the brachial artery, the wall of the artery collapses because more pressure is being exerted against the outside of the arterial wall than by the blood against the inside. If the pressure in the cuff is gradually decreased until it falls below systolic

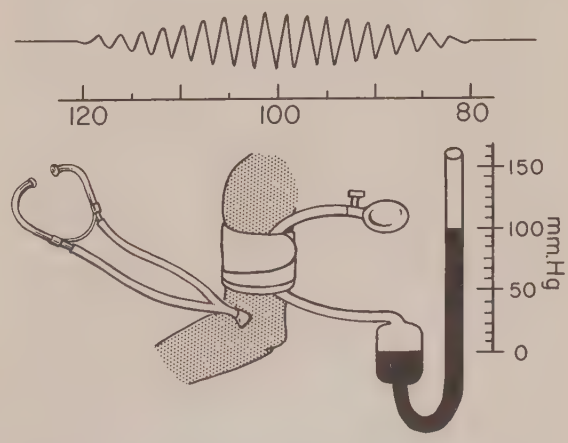


Figure 131. Measurement of systolic and diastolic pressures by the auscultatory method.

pressure, small spurts of blood begin to jet intermittently through the artery each time the arterial pressure rises to its systolic value. However, during diastole no blood flows through the artery. This intermittent flow of blood causes vibrations in the arteries of the lower arm that can be heard with a stethoscope. Therefore, to determine systolic pressure, one simply inflates the cuff until the pressure rises to a high value and then deflates it gradually until the intermittent sounds can be heard. At this instant the pressure recorded by the mercury manometer is a reasonably accurate measure of systolic pressure.

To determine diastolic pressure the cuff pressure is reduced still more. When the pressure falls below the diastolic value, blood then flows through the artery all of the time and no longer in jet-like bursts. As a result, the vibrations disappear, and sounds can no longer be heard with the stethoscope. The pressure recorded at this instant is a close estimate of the diastolic pressure. The intensity of the sounds heard through the stethoscope at different cuff pressure levels is shown graphically at the top of Figure 131.

Factors That Affect Pulse Pressure

The arterial *pulse pressure* is the pressure difference between the systolic and diastolic pressures. The higher the systolic pressure and the lower the diastolic pressure, the greater is the pulse pressure, but the more closely these two pressures approach each other, the less becomes the pulse pressure. The normal pulse pressure is about 40 mm. Hg. Two factors of major importance affect the pulse pressure in the arterial system: These are (1) stroke volume output of the heart, and (2) distensibility of the arterial system.

Stroke volume output. The stroke volume output is the amount of blood pumped by the heart with each beat. Normally, it is approximately 70 ml., but it can on occasion fall to as low as 10 to 20 ml., or can rise to as high as 160 ml. Obviously, if a small quantity of blood is pumped into the arterial system with each beat, the pressure will not rise and fall greatly during each cardiac cycle. But if the stroke volume output is great, the pressure will rise and fall tremendously, thereby giving an extremely high pulse pressure. In other words, the pulse pressure is roughly proportional to the stroke volume output.

The slower the rate of the heart, the greater must be the stroke volume output in order to maintain adequate blood flow through the body. Very slow heart beats are quite often found in well trained athletes, the slowness actually indicating efficient hearts. However, the large stroke volume output is accompanied by a high pulse pressure. On the other hand, persons with fever, toxicity, or weakened hearts very frequently have fast heart beats with accompanying small stroke volume outputs. The total amount of blood pumped per minute may be the same as that pumped by the athlete's heart, but because of the small stroke volume output the pulse pressure is very slight.

Distensibility of the arterial system. The more distensible the arteries, the greater the quantity of blood that can be compressed into the arterial system by a given amount of pressure. Therefore, each beat of the heart causes far less pressure rise in very distensible arteries than in non-distensible arteries.

Conditions that cause abnormal pulse pressure. In old age, the distensibility of the arterial system decreases tremendously because of arteriosclerotic changes in the walls of the vessels. *Arteriosclerosis* is usually the end result of atherosclerosis, which was described in the previous chapter, and arteriosclerosis actually means "hardening of the arterial walls." In many older persons who have arteriosclerosis, the walls even become calcified, bone-hard tubes. As a result, the arterial system then cannot stretch adequately during systole and cannot relax during diastole. Pulses of blood entering or leaving the arterial tree cause tremendous changes in pressure, sometimes causing a pulse pressure of 100 mm. Hg or more. A pulse pressure record from an arteriosclerotic patient is shown in Figure 129.

In *aortic regurgitation* the aortic valve has been partially destroyed. Blood then is pumped out of the heart into the arterial system, but during diastole much of the blood flows backward into the relaxed left ventricle, thus causing the diastolic pressure to fall very low—often to zero, as shown by the third curve in Figure 129. As a result, the pulse pressure becomes the greatest possible, sometimes as great as 160 mm. Hg.

In *patent ductus arteriosus,* which was described in Chapter 14, blood flows continually through the ductus from the aorta into the pulmonary artery and then through the lungs. This means that blood pumped by the left ventricle has an additional route by which it can leave the systemic arteries during diastole. The amount of blood that runs off from the arteries during diastole may be two to three times normal, which allows the pressure to fall to a very low diastolic value. Then, the extra blood flowing through the lungs returns to the left ventricle and increases the stroke volume output also to two to three times normal, causing the systolic pressure to rise tremendously. Thus, the pulse pressure often becomes two or more times normal, again causing a pressure pulse contour of the type illustrated by the third curve of Figure 129.

Transmission of Pressure Pulses to the Smaller Vessels

When the blood pumped by each beat of the heart is suddenly thrust into the aorta, the resulting increase in pressure at first distends only the portion of the aorta adjacent to the heart, as shown in Figure 132. A short period of time is required for some of this blood to push forward along the arterial tree and build up the pressure further peripherally. This movement of the pressure along the arteries is called *transmission of the pressure pulse.*

Transmission of the pressure pulse is analogous to the movement of a wave on the surface of a pond. When a pebble is thrown into the middle of the pond, a wave travels in every direction. When it reaches the shore, water rolls onto the bank and then back into the pond, causing the wave to turn around and go backward. The same happens in the arteries, for the pressure wave travels along the arteries away from the heart and upon reaching the very small arteries is often reflected back into the large arteries. Indeed, because of reflection, the contour of the pressure wave is sometimes quite different in the more peripheral arteries than in the base of the aorta. This is illustrated in Figure 133, which shows the normal pulse pressure contour in the proximal aorta and then a considerably different contour in the femoral artery as the result of a reflected wave. Also, it will be noted that the initial rise in the pressure contour in the femoral artery begins approximately 0.2 second after the beginning of the pressure rise in the aorta. This time lag is caused by the time required for transmission of the pressure wave from the aorta to the femoral artery.

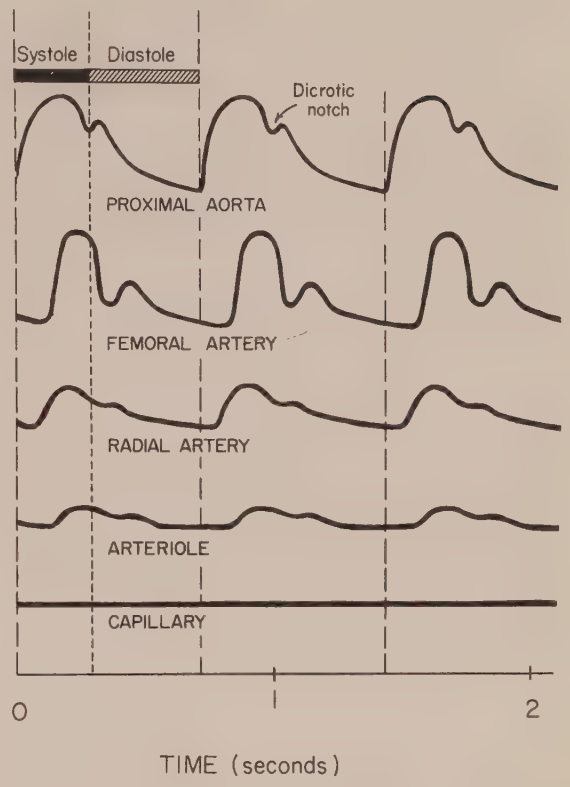


Figure 133. Changes in the pulse pressure contour as the pulse wave travels toward the smaller vessels.

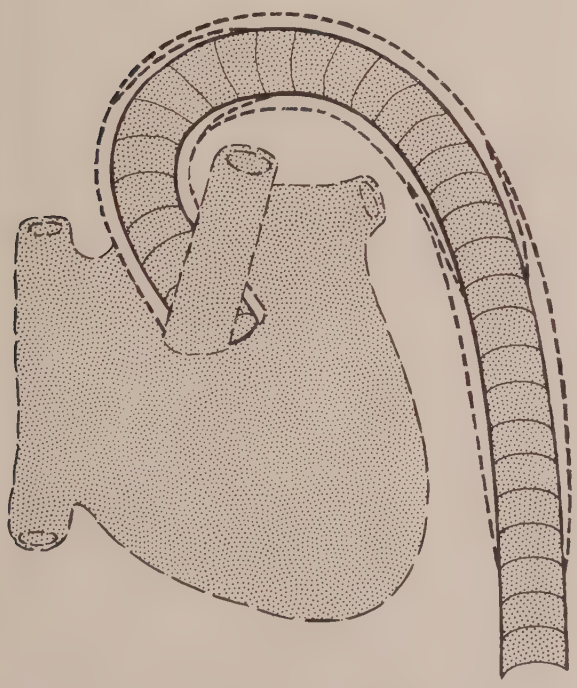


Figure 132. Progressive movement of the pressure pulse wave along the aorta.

Damping of the pressure pulses. Except when the pressure wave is reflected, the pulse pressure becomes less and less as the wave spreads toward the smaller vessels. This is called *damping* of the pulsations, and it results from two effects in the blood vessels: First, the resistance to blood flow in the artery not only impedes the flow of blood but also impedes the transmission of the pulse wave. Second, the walls of the vessels are not entirely elastic and, therefore, exert a "viscous" effect on the transmission of the pulse wave. That is, as the pressure attempts to distend and contract the vessels, the viscous nature of the vascular walls opposes the changes in pressure. Figure 133 illustrates the damping of the pulse wave as it spreads peripherally, showing especially the complete absence or almost complete absence of pulsations in the capillary.

PRESSURES AT DIFFERENT POINTS IN THE SYSTEMIC CIRCULATION

Figure 134 shows the pressures in the different vessels from the aorta to the venae cavae. From this chart it is obvious that about one half of the entire fall in pressure is in the arterioles and about one fourth is in the capillaries. The reason for this is that about one half of the resistance to blood flow is in the arterioles and about one fourth is in the capillaries.

In the normal person, the mean pressure in the aorta is approximately 100 mm. Hg, and it falls to about 85 mm. Hg at the end of the small arteries. The pressure then falls another 55 mm. Hg in the arterioles down to 30 mm. Hg at the beginning of the capillaries. At the end of the capillaries the pressure normally is about 10 mm. Hg and it gradually decreases in the venules, small veins, and large veins until it reaches 0 mm. Hg in the right atrium.

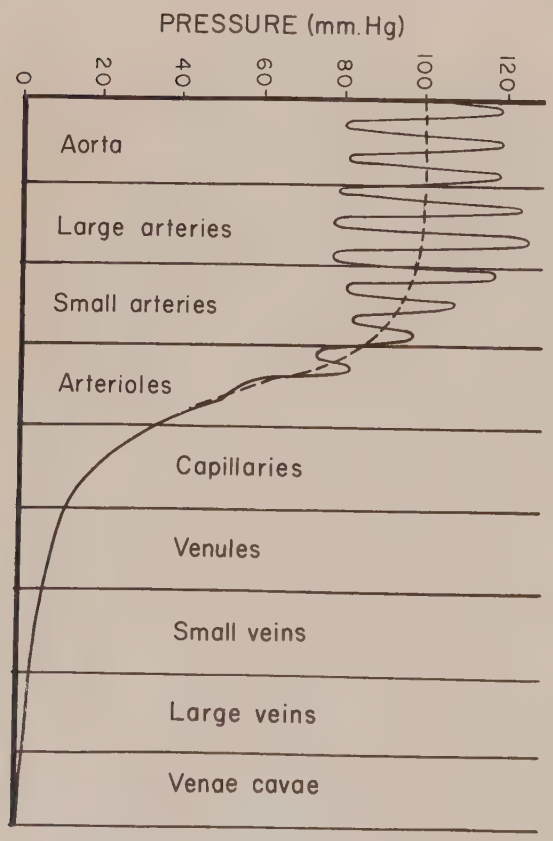


Figure 134. Pressures in different parts of the systemic circulation.

Effect of the Degree of Arteriolar Constriction on Pressures at Different Points in the Circulation

In the previous chapter it was pointed out that the arterioles are abundantly supplied with nerves that can cause them to contract and relax with ease. Also, the phenomenon of autoregulation can make the arterioles contract and relax in response to local conditions in the tissues. For these reasons, arteriolar resistance can change tremendously from time to time. It is important to understand how these changes affect pressures in the various portions of the circulation.

When the arterioles constrict totally, blood cannot flow into or out of the arterioles. As a result, the pressure in the small arteries rises to equal that in the aorta, and the pressure in the capillaries falls to equal that in the large veins.

When the arterioles relax and allow blood to flow with great ease, the capillary pressure rises to approach the pressure in the small arteries. This effect is especially important because a rise in capillary pressure can cause very rapid transfer of fluid into the tissues, and, conversely, a fall in the capillary pressure can cause rapid flow of fluid from the tissues back into the capillaries, as was explained in Chapter 6. Therefore, capillary dynamics are dependent to a great extent upon regulation of capillary pressure by the arterioles.

MEAN ARTERIAL PRESSURE

The *mean arterial pressure* is the arterial pressure averaged during a complete pressure pulse cycle. The mean arterial pressure is not always equal to the average of systolic and diastolic pressures, for during each pressure cycle the pressure usually remains at systolic levels for a shorter time than at diastolic levels. Therefore, when the pressures at all stages of the pressure cycle are averaged, the resulting mean pressure is usually slightly nearer diastolic pressure than systolic pressure.

So far as the circulatory system is concerned, *the mean arterial pressure is much more important than is either systolic or diastolic pressure,* because it is the mean pressure that determines the average rate at which blood will flow through the systemic vessels. For this reason, in most physiologic experiments it is not necessary to record the pressure changes throughout the pressure cycle, but instead it is necessary simply to record the mean arterial pressure. This can be accomplished admirably using the mercury manometer, which was illustrated in Figure 111 of

Chapter 13. Mercury has so much inertia that it cannot move rapidly enough to raise the recording arm to the systolic level during systole, and it will not fall enough to register diastolic pressure during diastole. Instead, it maintains a relatively even level with some minor pulsations during the beat of the heart; nevertheless, essentially, it records mean arterial pressure rather than the pulse contour.

Relationship of Mean Arterial Pressure to Total Peripheral Resistance and Cardiac Output

Recalling once again the basic formula given in Chapter 14 for the relationship of blood flow to pressure and resistance, the effect on the arterial pressure of vascular resistance and of the amount of blood pumped per minute is given by the following formula:

$$\text{Arterial pressure} = \text{Cardiac output} \times \text{Total peripheral resistance}$$

The *cardiac output* is the rate at which blood is pumped by the heart, and the *total peripheral resistance* is the total resistance of all the vessels in the systemic circulation, from the origin of the aorta back to the veins entering the right atrium. This means, then, that either generalized constriction of all the blood vessels or an increase in cardiac output will raise the arterial pressure. Conversely, dilatation of the vessels or a decrease in cardiac output will lower the pressure.

Regulation of Mean Arterial Pressure

Under resting conditions, the mean arterial pressure is normally regulated at a value of almost exactly 100 mm. Hg. The body has at least four separate types of arterial pressure regulatory systems which are responsible for maintaining this normal pressure. These are: (1) nervous mechanisms that regulate arterial pressure by controlling the degree of arteriolar constriction, (2) a capillary fluid shift mechanism which regulates arterial pressure by altering the blood volume, (3) the kidney excretory mechanism which also regulates arterial pressure by altering the blood volume, and (4) hormonal mechanisms that regulate either the blood volume or the degree of arteriolar constriction.

Nervous regulation of arterial pressure. In the previous chapter it was pointed out that almost all of the blood vessels of the body are supplied with sympathetic nerve fibers which, when stimulated, cause most of the blood vessels to constrict. Constriction of the vessels obviously impedes the flow of blood and, therefore, increases the total peripheral resistance, thereby increasing the arterial pressure, a fact that one can understand by referring to the preceding formula.

The nervous control of the circulation is illustrated in Figure 135. The insert to the left shows the medullary portion of the brain stem in which is located the so-called *vasomotor center* for control of (a) the degree of vasoconstriction of the blood vessels, (b) the degree of vasodilatation, and (c) the heart rate (cardioacceleration and cardioinhibition).

The vasomotor center controls the circulation mainly through the sympathetic nervous system, which is illustrated to the right in Figure 135. Nerve impulses are transmitted down the spinal cord into the sympathetic chains and then to the heart and blood vessels. These impulses increase the *rate of the heart* and its *strength of contraction*, both of which tend to increase the arterial pressure. They also cause *vasoconstriction* of the blood vessels in most parts of the body, which increases the total peripheral resistance and, therefore, increases the arterial pressure.

The vasomotor system also helps to control the circulation through the vagus nerves that go to the heart. When the sympathetics are stimulated, the parasympathetic fibers in the vagi are inhibited, and, conversely, when the sympathetic fibers are inhibited, the parasympathetic fibers are excited. Parasympathetic stimulation has almost the opposite effect on the heart to that of sympathetic stimulation; that is, it decreases the heart's activity. Except for this effect on the heart, the parasympathetics play almost no other role in control of the circulation.

VASOMOTOR TONE. Even normally, the sympathetic nervous system transmits impulses at a slow but continuous rate to the heart and vascular system, thus maintaining a moderate degree of vasoconstriction in the blood vessels. The sympathetic system decreases the arterial pressure by simply decreasing the number of impulses transmitted to the circulation, which allows the arteriolar walls to relax, thus allowing the arterioles to dilate and reduce the pressure. Conversely, it increases the arterial pressure by increasing the number of impulses above those normally transmitted, thus further constricting the arterioles.

Figure 136 demonstrates the importance of vasomotor tone in the regulation of arterial pressure. At the beginning of this record, the mean arterial pressure was 100 mm. Hg; then total spinal anesthesia was instituted by injecting a local anesthetic all the way up the spinal canal. This blocked all nerves coming from the spinal

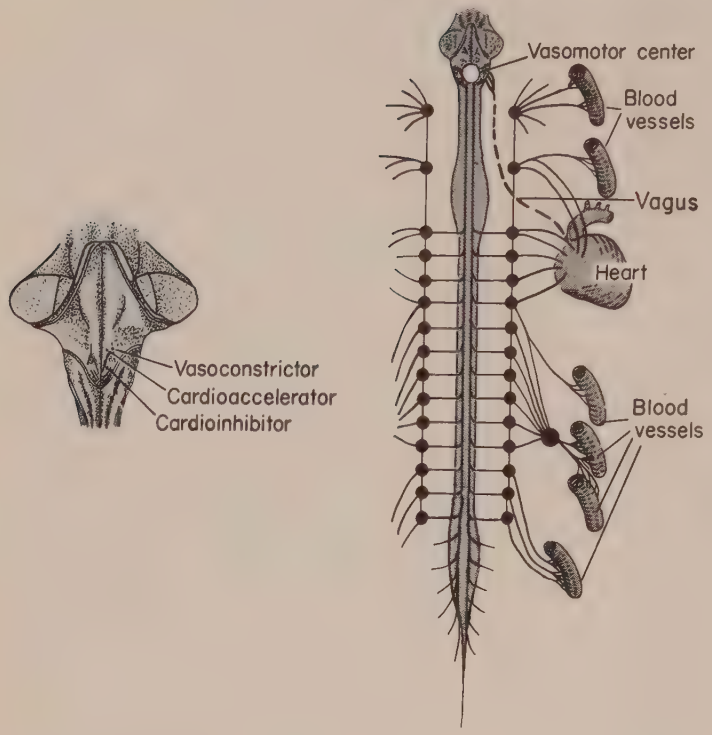


Figure 135. The vasomotor center, and its connections with the circulatory system through the sympathetic nervous system.

cord, including the sympathetic nerves. Note that the arterial pressure immediately fell to 50 mm. Hg, which is the pressure in the normal circulatory system when there is no vasomotor tone. After a few moments, a small amount of norepinephrine (the substance secreted at the sympathetic nerve endings) was injected into the blood. This caused vasoconstriction throughout the body and thereby immediately elevated the arterial pressure up to 150 mm. Hg. Yet, when the norepinephrine wore off after another few minutes, the arterial pressure returned to its basal level. Thus, we see from this experiment that arterial pressure can be reduced far below normal by decreasing the degree of vasomotor tone, and can be increased far above normal by increasing the tone.

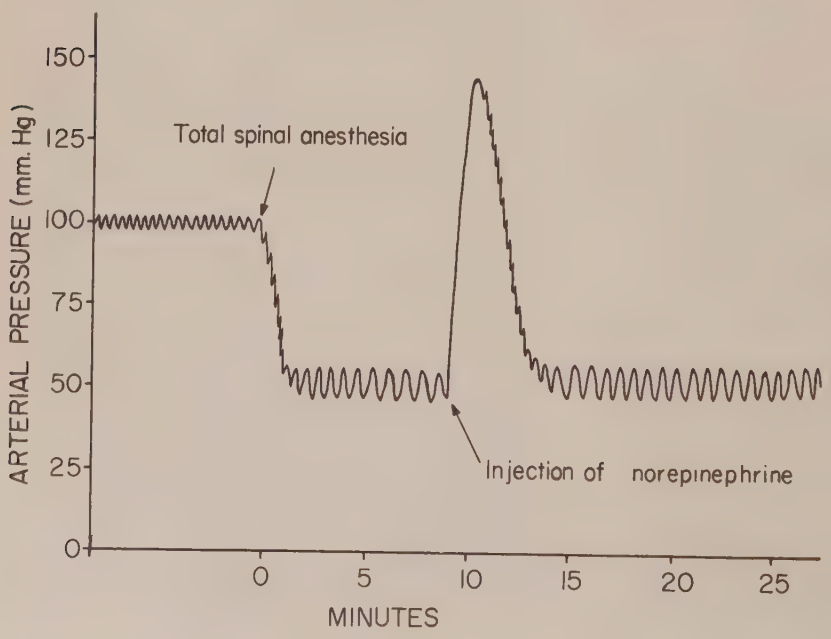


Figure 136. Effect on arterial pressure of greatly decreased vasomotor tone caused by spinal anesthesia, and of greatly increased vasomotor tone caused by injection of norepinephrine.

Factors That Affect Activity of the Vasomotor Center

Effect of carbon dioxide. One of the most powerful stimuli affecting the activity of the vasomotor center is carbon dioxide in the blood. When its concentration rises above normal, the vasomotor center becomes greatly excited, sending far more than normal numbers of impulses through the sympathetics to the blood vessels. As a result, the arterial pressure rises. A very large increase in carbon dioxide concentration can make the mean arterial pressure rise to as high as 260 mm. Hg, which is as high a pressure as the normal heart can pump.

This effect of carbon dioxide on arterial pressure helps to prevent malnutrition of the tissues, for, ordinarily, a high concentration of carbon dioxide in the blood is a result of very rapid metabolism in the tissues. The carbon dioxide excites the vasomotor center, the arterial pressure rises, this increases the rate of blood flow through the tissues, and the enhanced flow carries increased quantities of nutrients to the tissues.

Effect of medullary ischemia. The term *ischemia* means poor blood supply. When the vasomotor center becomes ischemic it becomes greatly excited, thereby elevating the arterial pressure. This effect is illustrated by the experiment of Figure 137, which shows fluid being injected under pressure into the space around the brain. The pressure against the brain compresses the arteries that supply the medulla, and the resultant medullary ischemia excites the vasomotor center, thereby elevating the arterial pressure.

The medullary ischemia mechanism normally protects the brain from being injured because of insufficient blood flow. This mechanism is equally as powerful as the effect of carbon dioxide, for it, too, can increase the mean arterial pressure to as high as 260 mm. Hg.

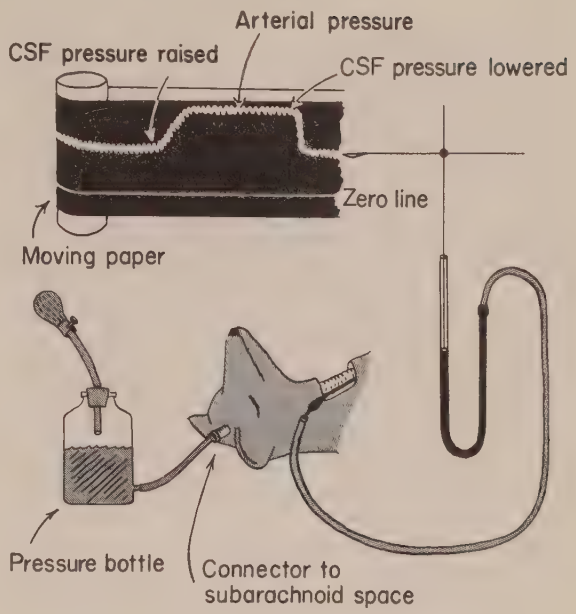


Figure 137. Elevation of arterial pressure caused by increased cerebrospinal fluid pressure.

The Pressoreceptor System

The pressure regulatory system that has been studied more than any other is the *pressoreceptor system* illustrated in Figure 138. In the arch of the aorta and in the internal carotid arteries of the neck are many small nerve receptors that detect the degree of stretch caused in the walls of the arteries by the pressure. These receptors are called *pressoreceptors.* The number of impulses transmitted from the pressoreceptors increases as the arterial pressure rises. On passing to the brain, the impulses inhibit the vasomotor center and thereby lower the arterial pressure back toward normal. On the other hand, when the pressure in the arteries falls too low, the pressoreceptors lose their stimulation so that the vasomotor center now becomes excessively excited, elevating the arterial pressure back to a higher level.

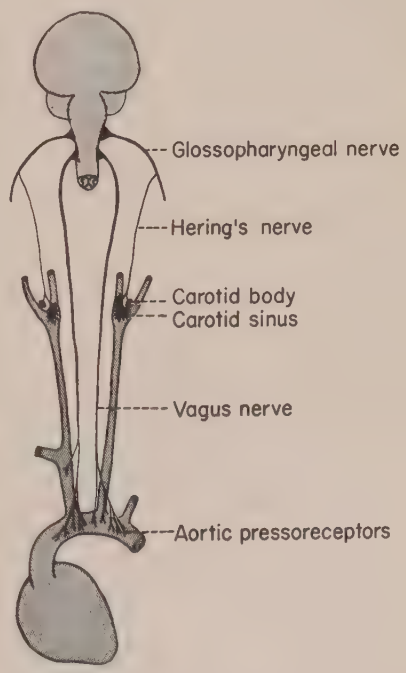


Figure 138. The pressoreceptors and their connections with the vasomotor center.

One can readily see that the pressoreceptor system opposes either a rise or a fall in pressure. For this reason, it is sometimes called a *moderator system* or a *buffer system.* As an example of this system's effectiveness, if 300 ml. of blood is injected into a person's circulation over a period of about 30 seconds, the arterial pressure will normally rise no more than 20 mm. Hg. However, injection of the same quantity of blood into the circulation of someone whose pressoreceptors have been inactivated will cause the pressure to rise about three times this much, or 60 mm. Hg.

The pressure in the arteries of the upper body tends to fall when one stands because of *pooling* of blood in the lower body, but the falling pressure automatically excites the pressoreceptor reflex, returning the pressure in the upper body almost to normal. One of the values of this constancy of pressure in the upper part of the body is that it helps to maintain an adequate blood supply to the brain regardless of the position of the body.

The Capillary Fluid Shift Mechanism for Pressure Regulation

Because of the rapidity with which the various nervous control systems can return arterial pressure toward normal when it becomes abnormal, it is very tempting to explain arterial pressure regulation entirely on the basis of nervous mechanisms alone. However, the nervous mechanisms have a major failing that prevents them from continuing to control arterial pressure over a long period of time. This failing is that the pressure receptors *adapt,* which means that they gradually become insensitive to the change in pressure. Therefore, after a few days, the nervous regulatory mechanisms become ineffective. By that time, non-nervous mechanisms take over the pressure regulation. One of these is the *capillary fluid shift mechanism,* which is particularly important in helping to regulate arterial pressure when the blood volume tends to become either too little or too great.

An increase in blood volume such as might occur following a transfusion increases the pressures in all parts of the systemic circulation, but over a period of a few minutes, the pressures return back toward normal. Much of this decline in pressure is caused by shift of fluid out of the blood through the capillary membranes into the interstitial spaces, thereby decreasing the blood volume back toward normal.

Conversely, when a person hemorrhages severely, fluid shifts out of the interstitial spaces into the circulation, increasing the blood volume back toward normal and returning the pressure also back toward normal.

The mechanism of the capillary fluid shift is simply the following: Too much blood volume increases the pressures in the capillaries as well as in the arteries. Therefore, in accordance with the law of the capillaries, which was explained in Chapter 6, the increased capillary pressure causes fluid to leak out of the circulation into the interstitial spaces. Conversely, when blood is lost from the circulation, the capillary pressure falls, and, here again in accordance with the law of the capillaries, fluid is pulled by the colloid osmotic pressure of the blood from the interstitial spaces into the circulation.

The capillary fluid shift mechanism is much slower to regulate arterial pressure than are the nervous mechanisms, for it requires from 10 minutes to several hours to readjust the arterial pressure back toward normal.

Regulation of Arterial Pressure by the Kidneys

The kidneys, like the capillaries, can regulate arterial pressure by increasing or decreasing the blood volume. After a substantial loss of blood, for instance, the kidneys simply stop forming urine because the glomerular pressure falls too low for glomerular filtration to take place. The fluid and electrolytes taken in by mouth gradually accumulate in the body until the blood volume returns to normal, reestablishing normal arterial pressure and at the same time reestablishing normal glomerular dynamics so that the kidneys begin to form urine again.

Conversely, when the blood volume becomes too great, the urinary output increases. Gradually, over a period of hours, the blood volume decreases back toward normal, causing the arterial pressure also to return toward normal.

Function of "renin" in arterial pressure regulation. Many physiologists believe that the kidneys also have a hormonal effect in regulating arterial pressure. It is postulated that when the arterial pressure falls to a low value, diminished blood flow through the kidneys causes them to secrete into the blood a special hormone called *renin.* Renin in turn combines with one of the plasma proteins of the plasma to form a substance called angiotensin which has been postulated to (a) act directly on the blood vessels to cause vasoconstriction, which increases the arterial pressure, or (b) act on the adrenal glands to cause increased aldosterone secretion, which in turn causes the kidneys to retain salt and water, the effect of which is to increase arterial pressure, as is explained below. Unfortunately, however, neither of these mechanisms has yet been proved to be

one of the usual pressure regulating mechanisms.

Importance of renal regulation of arterial pressure. Despite the uncertainty of the method by which the kidneys regulate arterial pressure, these organs seem to be the most important of all parts of the body for long-term regulation of arterial pressure. This is shown very dramatically by the experiment of Figure 139. If clamps are placed on the arteries of both kidneys so that the renal blood supply is diminished to 10 to 20 per cent of normal, the arterial pressure begins to rise. In several weeks, the mean pressure may be twice its original value. The elevated pressure will then force an adequate blood flow through the kidneys despite the continued constriction of the arteries, and the kidneys will function almost normally again. When the constrictions are removed, the arterial pressure falls back to normal within another few days. It is obvious that the kidneys are capable of protecting themselves against diminished blood flow; that is, *diminished flow through the kidneys causes a general arterial pressure rise until the kidneys again receive an adequate blood supply.* On the other hand, if the blood flow through the kidneys becomes excessive, the arterial pressure falls until the renal flow returns to normal.

Hormonal Regulation of Arterial Pressure—The Effect of Aldosterone

At least one major hormonal system is also involved in the regulation of arterial pressure. This is the secretion of aldosterone by the *adrenal cortex.* The adrenal cortex is the outer shell of two small endocrine glands, the adrenals, located immediately above the two kidneys. These cortices secrete adrenocortical hormones, one of which, *aldosterone,* decreases the kidney output of salt and water. The amount of salt and water in the body in turn helps to determine the volumes of blood and interstitial fluids in the body.

The manner in which the aldosterone mechanism enters into blood pressure regulation is the following. When the arterial pressure falls very low, lack of adequate blood flow through the body's tissues, for reasons not yet understood, causes the adrenal cortices to secrete aldosterone. The aldosterone acts to build up the blood volume and return the arterial pressure to normal. Conversely, an elevated arterial pressure reverses this mechanism so that fluid volumes, and consequently the arterial pressure, decreases.

Regulation of Mean Arterial Pressure during Exercise

One of the most severe types of stress to which the circulation can be subjected is strenuous exercise, for a rapidly acting muscle needs a tremendous blood supply. One of the means for increasing blood flow to the muscles during exercise is to elevate the arterial pressure. The mechanisms causing this are: First, increased muscular metabolism during exercise increases the concentration of carbon dioxide in the blood, and this in turn excites the vasomotor center to elevate the arterial pressure. Second, lactic acid and perhaps other metabolites are released from the exercising muscles; these also circulate through the blood to the vasomotor center and excite it, elevating the pressure. Finally, the vasomotor center is stimu-

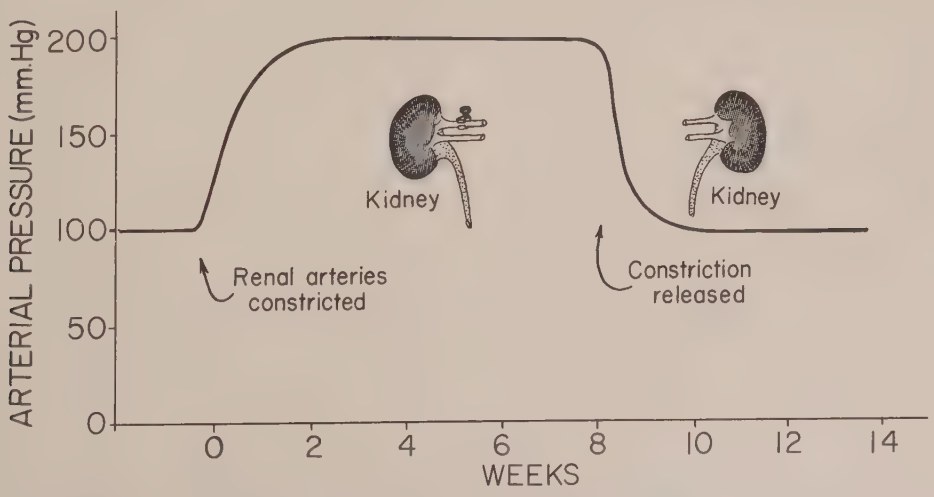


Figure 139. Effect on arterial pressure of constricting the renal arteries and weeks later releasing the constriction.

lated by nerve impulses generated in the motor area of the brain. Therefore, at the same time that the motor area causes the muscles to contract it also excites the vasomotor center and thereby elevates the arterial pressure.

HYPERTENSION

Hypertension means high arterial pressure, and it occurs in approximately one out of five persons. Hypertension can cause rupture of the blood vessels in the brain, in the kidney, or in other vital organs. Also, it can place excessive strain on the heart and cause it to fail. For these reasons one of the most important research problems in physiology is to determine the causes of hypertension. Some of the known causes are the following:

Hypertension Caused by Abnormal Function of the Pressure Regulatory Mechanisms

Renal hypertension. Any factor that damages the kidneys can cause *renal hypertension.* For instance, destruction of one whole kidney and two thirds of the second kidney causes the arterial pressure to rise, the greater the degree of destruction the greater the elevation of pressure. Diseases of the kidneys such as kidney infections, sclerosis of the renal blood vessels, kidney inflammation, and so forth can all elevate the pressure.

A particularly interesting type of renal hypertension occurs when the aorta is occluded above the kidneys, which occurs frequently before birth, in which case a person goes through life with an occluded aorta, a condition called *coarctation of the aorta.* Blood finds its way through many smaller arteries from the upper segment of the aorta into the lower segment, but the resistance to flow through these small arteries is so great that the arterial pressure in the lower segment is much lower than that in the upper segment. The insufficient blood flow to the kidneys from the lower aorta causes the pressures throughout the body to rise until the pressure in the lower aorta returns to normal. Obviously, by this time, the pressure in the upper aorta will have risen to a very high level. Thus, the kidneys automatically maintain a normal arterial pressure at the kidney level even at the expense of the arterial pressure in the upper part of the body.

MECHANISM OF RENAL HYPERTENSION. The precise mechanism of renal hypertension is not known, but there are two prevalent theories. A theory that has been taught for the past 35 years is based on the renin mechanism for arterial pressure regulation. It is claimed that a damaged kidney or a kidney with a poor blood supply secretes very large quantities of renin which, as explained previously, causes either vasoconstriction or increased aldosterone secretion which eventuates in hypertension.

A second theory to explain the mechanism of renal hypertension is based on alterations in fluid volumes of the body when renal function becomes abnormal. Obviously, if renal output becomes reduced as a result of renal disease or poor blood supply, the retention of salt and water would be expected to increase both the interstitial and blood volumes. Experiments have recently demonstrated that removal of large portions of the kidney do indeed increase both of these volumes considerably. It is postulated that these increased volumes increase the arterial pressure. Presumably, the nervous pressure regulating mechanisms prevent the increased pressure at first, but after the nervous mechanisms have adapted for several days, the arterial pressure, being no longer inhibited by the nervous control system, rises to hypertensive levels.

Hormonal hypertension. Occasionally the adrenal cortices become excessively active, either because of a secreting tumor in one of the adrenal glands or because of excessive stimulation of the adrenals by the anterior pituitary gland. In any event the increased production of aldosterone causes the kidneys to retain excessive quantities of water and salt. The fluid volumes throughout the body increase, and the arterial pressure often rises to as high as two times its normal pressure.

A second type of hormonal hypertension is that caused by a *pheochromocytoma.* This is a tumor of the adrenal *medulla,* which is a portion of the adrenal gland entirely distinct from the adrenal cortex. Because the medulla is part of the sympathetic nervous system, the pheochromocytoma secretes large quantities of epinephrine and norepinephrine, the hormones normally secreted by sympathetic nerve endings. These hormones circulate through the blood to all the vessels and promote intense vasoconstriction. As a result, the mean arterial pressure may rise to as high as 200 mm. Hg or higher. The pheochromocytoma usually secretes its hormones only when stimulated by the sympathetic nervous system. This means that when a person who has this type of tumor becomes excessively excited or has some other cause for stimulation of the sympathetic system, the secretion of epinephrine and norepinephrine may be tremendous, causing the blood pressure to become very high, while between periods of excitement the blood pressure may be essentially normal.

Neurogenic hypertension. Hypertension often results from abnormalities in the brain. For instance, sudden blockage of one of the arteries to vital portions of the vasomotor center sometimes causes hypertension that lasts for life. Also, experiments in animals have indicated that destruction of certain areas in the hypothalamus can cause prolonged hypertension. These abnormalities cause the hypertension by continually exciting the vasomotor center.

Many psychiatrists believe, though it has not yet been proved, that overexcitation of certain of the conscious regions of the brain in neurotic states such as anxiety, worry, and so forth, can excite the vasomotor center and permanently elevate the arterial pressure. Finally, it has been observed many times that excessive pressure in the cranial vault, resulting from vascular hemorrhage, brain tumors, or other causes can elicit the medullary ischemic reflex, thereby causing high arterial pressure.

Essential Hypertension

The types of hypertension described above have known causes. Yet in approximately 90 per cent of all hypertensive persons the cause is unknown, and they are said to have *essential hypertension,* the word essential meaning simply "of unknown cause." Many efforts have been made to prove that essential hypertension is caused by a kidney abnormality, a glandular abnormality, or excessive activity of the vasomotor center. Authentic proof that any of these factors is involved in essential hypertension is yet lacking, which means that the physiologist is still faced with discovering the cause of high blood pressure in most patients. A possible cause of essential hypertension might be a genetic abnormality of the arterioles, causing them always to constrict to excessive degrees. Indeed, it is a known fact that essential hypertension exhibits a very strong hereditary tendency. If one person should inherit excessively active arterioles and another more relaxed arterioles, the first would have hypertension and the second would be normal. Also, it has been suggested that the person with essential hypertension might have a genetic deficiency of arterioles, which obviously would impede the flow of blood and cause hypertension. Thus far, researchers have not found means of exploring these possibilities, though they must be included among the others as possible causes of essential hypertension.

REFERENCES

Freis, E. D.: Hemodynamics of hypertension. *Physiol. Rev.* 40:27, 1960.

Guyton, A. C., and Harris, J. W.: Pressoreceptor-autonomic oscillation: a probable cause of vasomotor waves. *Am. J. Physiol.* 165:158, 1951.

Heymans, C., and Neil, E.: *Reflexogenic Areas of the Cardiovascular System.* Boston, Little, Brown and Company, 1958.

Hoff, E. C., Kell, J. F., and Carrol, M. N.: Effects of cortical stimulation and lesions on cardiovascular function. *Physiol. Rev.* 43:68, 1963.

Page, I. H.: *Arterial Hypertension.* Philadelphia, F. A. Davis Company, 1960.

Page, I. H., and Bumpas, F. M.: Angiotensin. *Physiol. Rev.* 41:331, 1961.

Tobian, L.: Interrelationship of electrolytes, juxtaglomerular cells and hypertension. *Physiol. Rev.* 40:280, 1960.

CARDIAC OUTPUT, VENOUS PRESSURE, CARDIAC FAILURE, AND SHOCK

16

CARDIAC OUTPUT

Cardiac output is the rate at which the heart pumps blood, and ordinarily it is expressed in liters or milliliters per minute. Because the function of the circulatory system is to supply adequate nutrition to the tissues, the subject of cardiac output is one of the most important of all phases of physiology.

Normal Values of Cardiac Output

The average cardiac output in a person lying down and in a state of complete rest is about 5 liters per minute. If he walks, it rises to perhaps 7.5 liters per minute. If he performs strenuous exercise, it might rise to as high as 25 liters per minute in the normal person or to as high as 35 liters per minute in the well trained athlete.

From the preceding figures, one can see that the cardiac output varies in proportion to the degree of activity of the person. The ability to regulate the cardiac output in accord with the needs of the body is one of the most remarkable feats of the circulation. Much of the present chapter is devoted to discussing this regulation.

Regulation of Cardiac Output

The amount of blood that the heart can pump each minute is determined by two major factors: (1) the pumping effectiveness of the heart itself, and (2) the ease with which blood can return to the heart from the systemic circulation after it is pumped by the heart. To understand the regulation of cardiac output it is essential that one be familiar with the way in which each of these factors affects the circulation.

The pumping effectiveness of the heart. Normally the heart acts like a "sump" pump, pumping any blood that flows into it from the veins. If the amount flowing from the veins increases, the heart becomes stretched and automatically adjusts to accommodate the extra blood. This ability of the heart to pump any amount of blood, within limits, is called the *law of the heart,* which was explained in detail in Chapter 12. Figure 140 illustrates the principle pictorially and graphically, showing that as the heart becomes progressively stretched by greater and greater inflow of blood, the amount pumped increases up to the point at which the heart muscle finally becomes stretched beyond its physiologic limits.

EFFECT OF NERVOUS STIMULATION ON THE EFFECTIVENESS OF THE HEART AS A PUMP. Stimulation of the sympathetic nerves to the heart greatly enhances the effectiveness of the heart as a pump, while stimulation of the parasympathetic nerves (the vagi) greatly decreases the pumping effectiveness. Under resting conditions parasympathetic stimulation keeps the heart depressed, which probably prolongs the life of the heart. During exercise and during other circulatory states when the cardiac output needs to be enhanced very greatly, sympathetic stimulation takes over and makes the heart a very effective pump, about two-thirds more effective than the normal unstimulated heart.

Therefore, the pumping effectiveness of the heart can be increased in two separate ways: (1) by local adaptation of the heart to increased inflow of blood, the mechanism called the law of the heart and (2) by a shift from parasympathetic to sympathetic stimulation.

Flow of blood into the heart (venous return). Because the heart pumps whatever amount of blood enters its chambers, cardiac output normally is regulated mainly by the amount of blood that returns to the heart from the peripheral vessels. This flow of blood into the heart is called *venous return.* When a person is at rest, venous return averages 5 liters per minute. Therefore, this is the amount of blood that is pumped. However, if 40 liters per minute, a very abnormal amount, should attempt to return to the heart, the normal heart, even one strongly stimulated by the sympathetics, would be able to pump only

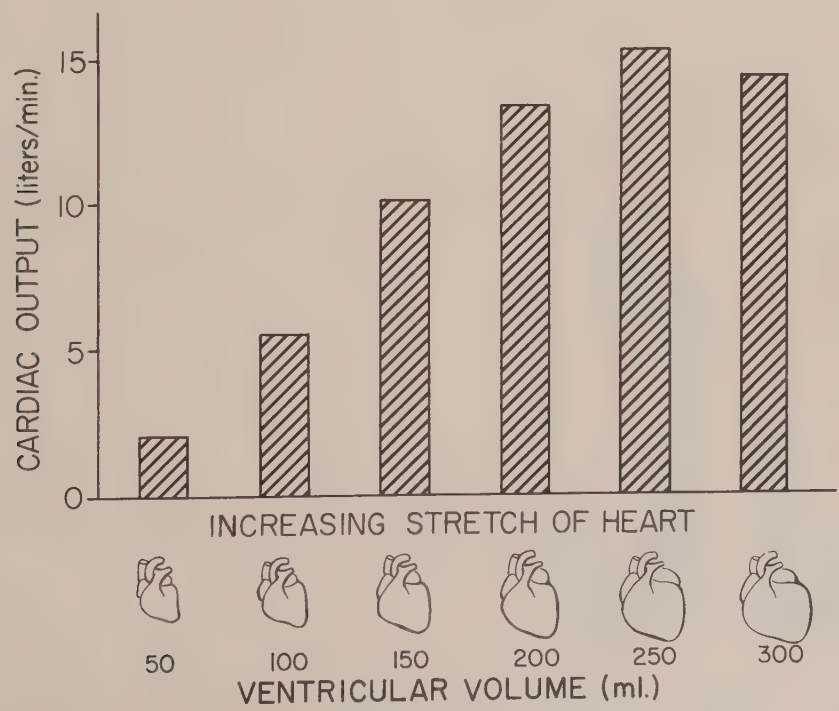


Figure 140. Effect of increased cardiac filling on the heart's ability to pump blood.

about 25 liters of this. Therefore, blood would dam up in the veins rather than being pumped into the arteries.

The important factors that determine the amount of venous return per minute are (1) the average pressure of the blood in all parts of the systemic circulation, which is called the *mean systemic pressure,* (2) the right atrial pressure, and (3) the resistance to blood flow through the systemic vessels.

EFFECT OF MEAN SYSTEMIC PRESSURE ON VENOUS RETURN. The *mean systemic pressure* is the average pressure of all the stored blood in the systemic circulation. Far more blood is stored in the veins than in the arteries. Therefore, the mean systemic pressure is more nearly equal to the venous pressure than to the arterial pressure. The normal mean systemic pressure is 7 mm. Hg. If the circulation becomes overfilled with blood or if the blood vessels contract tightly around the blood, the mean systemic pressure can rise to two to three times normal. On the other hand, loss of blood or loss of vasomotor tone causes the mean systemic pressure to fall far below normal.

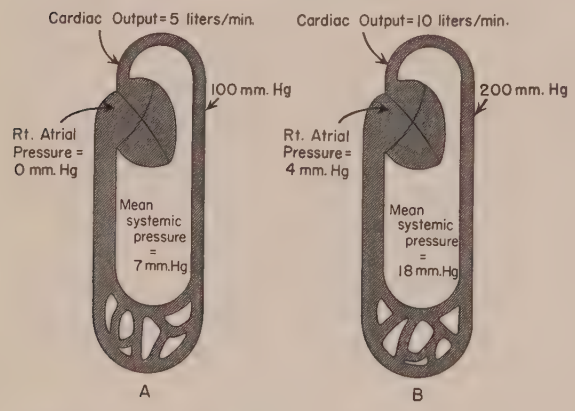


Figure 141. Effect of mean systemic pressure and right atrial pressure on venous return.

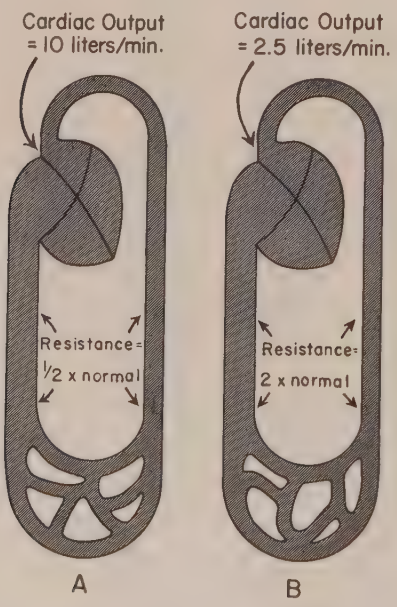


Figure 142. Effect of systemic vascular resistance on venous return.

Figure 141 illustrates the effect on cardiac output of changing the mean systemic pressure. Figure 141A shows a mean systemic pressure of 7 mm. Hg, in which case the average force pushing the stored blood in the systemic circulation toward the heart is 7 mm. Hg. Figure 141B shows a mean systemic pressure of 18 mm. Hg. In this instance the average force pushing the stored blood toward the heart is more than two times that in Figure 141A. Obviously, the venous return to the heart should be greatly increased. One sees from the figure that the cardiac output (which is equal to venous return) is twice as great in Figure 141B as in 141A.

EFFECT OF RIGHT ATRIAL PRESSURE ON VENOUS RETURN. The higher the right atrial pressure, the more difficult is it for blood to flow from the systemic circulation into the right atrium. Even a 3 to 4 mm. Hg increase in right atrial pressure, if it occurs without a concomittant rise in mean systemic pressure, will reduce the venous return to about one-half normal, thus illustrating the very great importance of this factor in determining how much blood will enter the heart each minute. Figure 141 shows this effect of right atrial pressure on venous return. In Figure 141A the right atrial pressure is zero, and venous return is proportional to the mean systemic pressure minus the right atrial pressure, which is equal to $7 - 0$, a total *driving force* of 7 mm. Hg. In Figure 141B the right atrial pressure has risen to $+4$ mm. Hg while the mean systemic pressure has risen to 18 mm. Hg. The difference between these two values, 14 mm. Hg, is the total driving force for moving blood from the systemic circulation toward the heart, a force called the *pressure gradient for venous return.* In this second case, the pressure gradient for venous return is exactly two times as great as the pressure gradient for venous return in Figure 141A. Therefore, the venous return and cardiac output also are exactly two times as great.

EFFECT OF VASCULAR RESISTANCE ON VENOUS RETURN. If all other factors remain constant, an increase in resistance to blood flow in the systemic circulation, particularly an increase in the resistance in the veins, reduces the amount of blood that returns to the heart, and this correspondingly reduces the cardiac output. Figure 142A shows

that if the resistance in all vessels of the body is decreased to one-half normal, the cardiac output will become two times normal, or 10 liters per minute. Figure 142B shows that if the resistance increases to two times normal, the cardiac output becomes only one-half normal, or 2.5 liters per minute.

Equilibrium between venous return and cardiac output. Obviously, the average venous return must equal cardiac output, though occasionally for only a few beats of the heart the two can be different from each other. When this occurs venous return and cardiac output automatically readjust to each other within a few seconds. Figure 143 illustrates the reasons for this readjustment. When the right atrial pressure is −4 mm. Hg, the normal venous return will be approximately 6 liters per minute, while the cardiac output will be only 2 liters per minute. Obviously, the amount of blood returning to the right atrium under these circumstances will be greater than the amount pumped out. As a result, the volume and the pressure in the right atrium will be increasing. When the right atrial pressure has risen to zero, the venous return will have fallen to 5 liters per minute and the cardiac output will have risen to this same value. Thus, the circulation automatically reequilibrates venous return and cardiac output when some abnormal circumstance temporarily makes venous return greater than cardiac output. Exactly the opposite effects occur when the cardiac output becomes greater than venous return. Usually only four to six beats of the heart are required for the reequilibration.

Effect of special factors on cardiac output.

TRANSFUSION OF BLOOD. Transfusion of blood into a person increases the volume of blood in the peripheral vessels and thereby increases cardiac output because of two very important effects. First, the increased volume *increases the mean systemic pressure* which forces an increased quantity of blood toward the heart. Second, the increased blood volume dilates the vessels and thereby decreases the resistance to blood flow. This also increases venous return.

CHANGES IN VASOMOTOR TONE. Increased tone of the blood vessels normally increases cardiac output. The reason for this is that increased tone contracts the blood reservoirs, thus forcing increased amounts of blood from the veins toward the heart. Therefore, sympathetic stimulation, which increases the vasomotor tone all through the body, increases the cardiac output. Administration of various vasoconstrictor drugs also usually increases the output.

EXERCISE. Exercise greatly increases the cardiac output for several reasons. At the onset of exercise the muscles become tense over the entire body, which compresses all the blood vessels instantaneously and thereby increases the mean systemic pressure from the normal value of 7 mm. Hg up to more than double this value. This immediately enhances the venous return to the

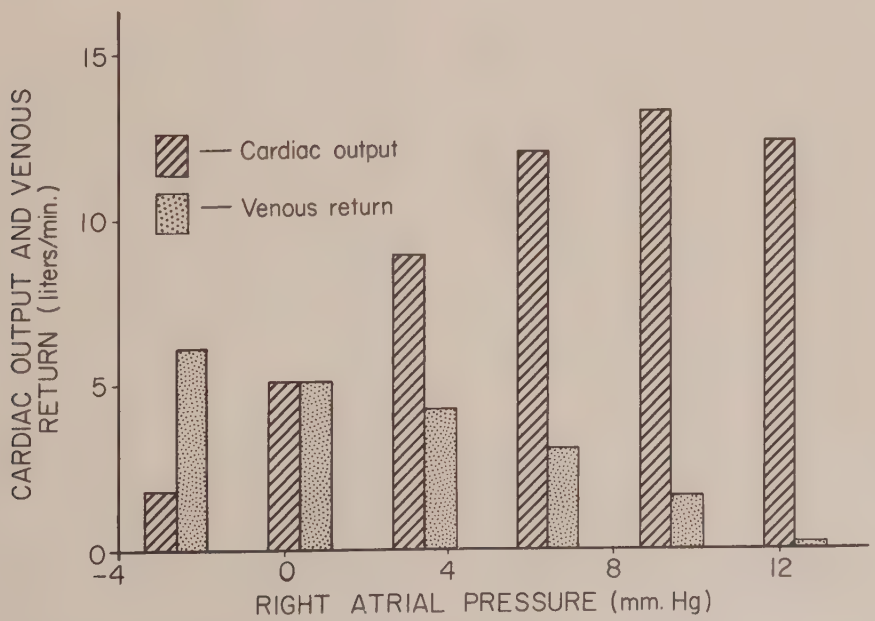


Figure 143. Equilibration of venous return and cardiac output at a right atrial pressure of 0, and lack of equilibration at all other right atrial pressures.

heart and causes the cardiac output to increase a moderate amount instantaneously.

Then, several other factors follow that increase the output still more. The sympathetic nervous system becomes strongly stimulated, causing the blood reservoirs to contract and thereby increase the mean systemic pressure still more. By the time the sympathetic nervous system has become fully excited, within 10 to 20 seconds, the mean systemic pressure will have risen to as much as four times normal, and there will have been a corresponding increase in cardiac output.

A final factor that increases the output is the effect of local *autoregulation* in the active muscles themselves. The high rate of muscular metabolism during exercise causes a relative deficiency of oxygen and of other nutrients in the muscles, which results in muscular vasodilatation, as was discussed in detail in Chapter 14. This in turn allows blood to flow very easily through the muscle vessels back to the heart. Thus, the total resistance to blood flow through the systemic circulation is greatly decreased, thereby promoting markedly increased venous return and correspondingly increased cardiac output. This change in vascular resistance in exercising muscles explains why cardiac output normally increases in proportion to a person's degree of activity.

The most intense degrees of exercise can increase cardiac output to as high as 20 to 25 liters per minute in a normal person and to as high as 30 to 35 liters per minute in the well trained athlete.

Summary of cardiac output regulation. Cardiac output is regulated, first, by the tendency for blood to return to the heart, and, second, by the effectiveness of the heart as a pump. Ordinarily, the normal heart is capable of pumping all the blood that attempts to return to it from the peripheral circulatory system. Therefore, it is frequently said that cardiac output is "regulated by the venous return." However, the amount of blood attempting to return to the heart often rises above that which the normal resting heart can pump. As a result, blood dams up in the veins, the venous pressure rises, and the heart is said to be *overloaded*, or said by some physiologists to be in *failure*.

In the preceding sections we have discussed many details of cardiac output regulation but not an overall theory of cardiac output regulation. The basic factor that determines the cardiac output is the need of the tissues for nutrition, and the nutrient that is most limited to the tissues is oxygen. Furthermore, experiments have shown that blood flow through most tissues of the body is directly proportional to the needs of the tissues for oxygen. Thus, each tissue regulates locally its own blood flow, and in doing so increases the venous return when the body's metabolism increases, which obviously increases the cardiac output. To state this another way, the accumulative need of the tissues for blood flow is the principal factor that regulates cardiac output.

Yet there are times when the normal unstimulated heart simply cannot pump to all the tissues sufficient blood flow to satisfy the accumulative need. Under these conditions, particularly in exercise, the sympathetic nervous system usually becomes very powerfully stimulated and the parasympathetic system becomes inhibited. These two effects work together to increase the heart rate, to increase the strength of the heart, and to constrict the blood reservoirs in the systemic circulation, all of which work together to increase the cardiac output still more.

Measurement of Cardiac Output

A *flowmeter* can be inserted into the arteries or veins of an experimental animal in such a manner that all the blood flowing through the heart will be recorded. Such procedures, however, are not very useful even in animal experimentation because of the extreme disruption of normal function resulting from the necessary surgery. Obviously, in the human being they could not be used at all. Therefore, for measuring cardiac output in human beings and in intact animals the following indirect procedures have been developed.

The dye method. A simple procedure for measuring cardiac output is to inject a small quantity of foreign substance, such as a brightly colored dye, rapidly into a vein. As the substance passes through the heart and lungs into the arterial system, its concentration in the arterial blood is recorded by a photoelectric "densitometer" or by taking periodic blood samples for the next 20 to 30 seconds and analyzing these for their content of dye. The more rapidly the blood flows through the heart, the more rapidly the dye appears in the arteries and the sooner it will disappear. From the recorded concentration curve of the dye as it passes through the arteries one can calculate the cardiac output quite accurately.

The Fick principle. Another indirect method for measuring cardiac output is based on the so-called Fick principle, which is explained by the example of Figure 144. The oxygen in a sample of venous blood is analyzed chemically to be 160 ml. of oxygen per liter of blood, and in the arterial blood, 200 ml. Thus, as each liter of blood passes through the lungs it picks up 40 ml. of oxygen. At the same time, using appropriate respiratory apparatus, the amount of oxygen absorbed from

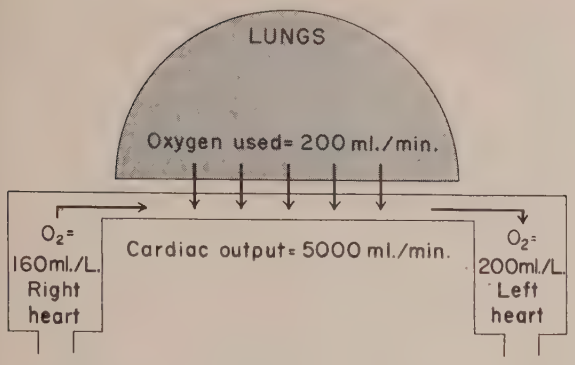


Figure 144. The Fick principle for indirectly measuring cardiac output.

the lungs is measured to be 200 ml. per minute. From these values, one can calculate the cardiac output in the following manner. If the amount of oxygen absorbed per minute is 200 ml. and the amount of oxygen picked up by each liter of blood flowing through the lungs is 40 ml., dividing 200 by 40, one can conclude that five 1-liter portions of blood must pass through the lungs each minute. Thus, the total cardiac output is calculated to be 5×1 liter or 5 liters per minute.

Expressing the Fick principle mathematically the following formula applies:

$$\text{Cardiac output in liters per minute} = \frac{\text{Total oxygen absorbed per minute}}{\text{Oxygen absorbed by each liter of blood}}$$

VENOUS PRESSURE

Right Atrial Pressure

The pressure in the venous system is determined mainly by the pressure in the right atrium. Normally, the right atrial pressure is approximately zero—that is, it is almost exactly equal to the pressure of the air surrounding the body. This does not mean, though, that the force distending the walls of the right atrium is zero, because the pressure in the thoracic cavity surrounding the heart is about 4 mm. Hg less than atmospheric pressure. This partial vacuum actually pulls the walls of the atrium outward so that blood is normally sucked from the veins into the atrium.

If the heart becomes weakened, blood begins to dam up in the right atrium, thereby causing the right atrial pressure to rise. Or, if excessive quantities of blood attempt to flow into the heart from the veins, the right atrial pressure rises. Therefore, the right atrial pressure, though normally almost exactly zero, often rises above zero in abnormal conditions.

Peripheral Venous Pressures

The pressure in a systemic vein is determined by five major factors: (1) the right atrial pressure, (2) the resistance to blood flow from the vein to the right atrium, (3) the rate of blood flow along the vein, (4) hydrostatic pressure, and (5) the venous pump. The effect of most of these factors is self-evident, but some of them need special explanation.

Since blood flows always toward the heart, the pressure in any peripheral vein of a person in the lying position must be as great as or greater than the pressure in the right atrium. If considerable resistance exists between the peripheral vein and the heart, and the flow of blood is of reasonable quantity, then the venous pressure will be considerably higher than right atrial pressure. Many of the veins are compressed where they pass abruptly over ribs, between muscles, between organs of the abdomen, or so forth. At these points the venous resistance is usually very great, impeding blood flow and increasing the pressure in the distal veins. On the average, the pressure in a vein of the arm or leg of a person lying down is approximately 6 to 8 mm. Hg.

Effect of hydrostatic pressure on peripheral venous pressure. Hydrostatic pressure is the pressure that results from the weight of the blood itself. Figure 145 illustrates the human being in a standing position, showing that for blood to flow from the lower veins up to the heart, considerable extra pressure must develop in these veins to drive the blood uphill. The weight of the blood from the level of the heart down to the bottom of the foot is great enough that when all other factors affecting venous pressure are non-operative, the venous pressure in the foot will be 90 mm. Hg.

The venous pump. To prevent the extremely high venous pressures that hydrostatic pressure can cause, the venous system is provided with a special mechanism for propelling blood toward the heart. This is called the *venous pump* or sometimes the *muscle pump*. All peripheral veins contain valves which allow blood to flow only toward the heart. Every time a muscle contracts or every time movement of any portion of a limb is effected in any other way, the blood in the veins is compressed. Since the valves prevent backward flow, the blood is always pushed toward the heart, in this way emptying the veins and decreasing the peripheral venous pressure.

The hydrostatic pressures shown in Figure 145 occur only when a person is standing completely still or when some disease condition has destroyed

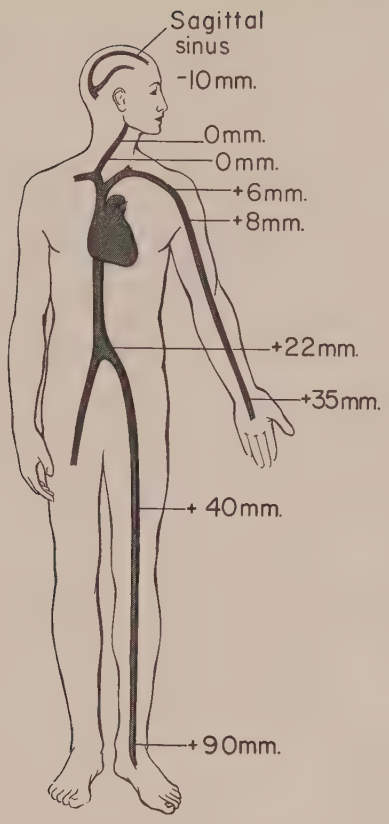


Figure 145. Hydrostatic pressures in various parts of the venous system of a person standing quietly so that the venous pump is inactive.

the valves of his veins. Ordinarily, the venous pump is so effective that pressures in the leg veins are only 15 to 30 mm. Hg. But when the valves are destroyed, such high pressures (80 to 90 mm. Hg) develop in the leg veins that they become progressively distended to diameters four and five times normal, causing the condition known as *varicose veins.*

Measurement of Venous Pressures

The most important of all venous pressures is the right atrial pressure, for one can usually tell from this how well the heart is pumping. Right atrial pressure is usually measured through a cardiac catheter in the manner illustrated in Figure 146A. The catheter is threaded into a vein of the arm, then upward through the veins of the shoulder, into the thorax, and finally into the right atrium. The pressure transmitted through the catheter is recorded by a water manometer located at the patient's side. This same technique can be used for recording pressures in other parts of the heart or pulmonary circulation simply by sliding the catheter through the tricuspid valve into the right ventricle, or on through the pulmonary valve into the pulmonary artery.

For measuring peripheral venous pressures, either a catheter or a needle can be inserted directly into a peripheral vein and the pressure measured as above. However, one can usually estimate peripheral venous pressures quite satisfactorily using the simple technique illustrated in Figure 146B of raising and lowering the arm above or below the level of the heart. When the veins are lower than the heart they become full of blood and stand out beneath the skin. Then, as the arm is raised, the veins normally collapse at a level approximately 9 cm. above the level of the heart. This means that the pressure in the arm veins, when the arm is at heart level, is about 9 cm. of water. If the veins do not collapse until they rise to a level of 20 cm. above the heart, then the peripheral venous pressure is approximately +20 cm. of water.

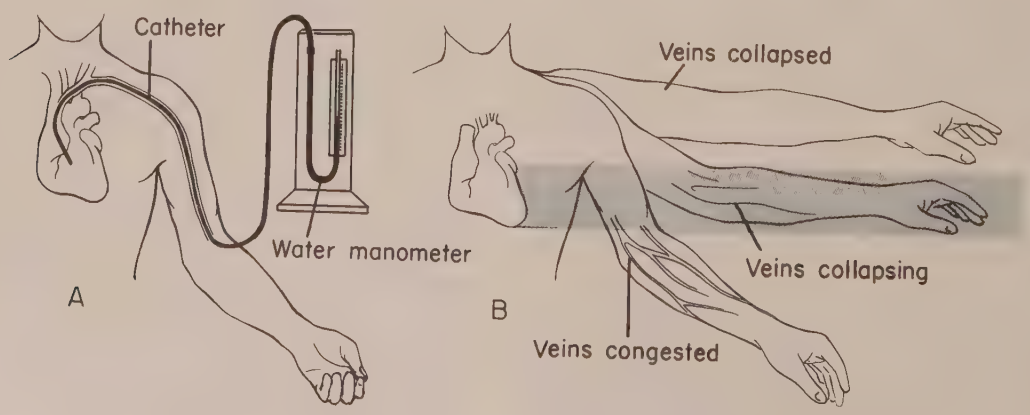


Figure 146. Measurement of (A) right atrial and (B) peripheral venous pressures.

CARDIAC FAILURE

The term *cardiac failure* means *depressed pumping effectiveness of the heart.* The most frequent cause of cardiac failure is actual damage to the heart itself by some disease process. For instance, the coronary arteries might become blocked because of arteriosclerosis, or sometimes the valves of the heart are destroyed by rheumatic heart disease. In either case, the effectiveness of the heart as a pump is decreased so that even the normal amount of blood is not pumped as well as normally.

Because the heart is actually two separate pumps, it is possible for one side of it to fail independently of the other. That is, the left ventricle can fail and dam blood in the pulmonary circulation, or the right ventricle can fail and dam blood in the systemic circulation.

Low cardiac output failure. A failing heart often fails to pump adequate quantities of blood to the tissues. This is called *low cardiac output failure.* It can be caused by weakness of any part of the heart or of the whole heart, because failure of any one part can often hinder satisfactory pumping.

Pulmonary congestion resulting from cardiac failure. When it is primarily the left side of the heart that is failing, the right heart continues to pump blood into the lungs with normal vigor while the left heart is unable to move the blood on into the systemic circulation. The resulting accumulation of blood in the lungs increases the pressures in all the pulmonary vessels and engorges them with blood. If the pulmonary capillary pressure becomes high enough, fluid will leak into the lung tissues, resulting in *pulmonary edema.* Indeed, in severe cases of left heart failure, fluid can even leak into the air spaces of the lungs, resulting in *hypostatic pneumonia.* Many patients who die from failure of the left heart do not die because of decreased cardiac output, but because of failure of the water-soaked lungs to aerate the blood.

Peripheral congestion and edema resulting from cardiac failure. If the right side of the heart fails or if the entire heart fails, the right atrial pressure rises, causing much of the blood attempting to return to the heart to be dammed in the peripheral veins. As a result, the pressures throughout the entire venous system rise. The veins of the neck may be greatly distended, as illustrated in Figure 147, and the venous reservoirs such as the liver and spleen may become engorged with blood. Also, the capillary pressure throughout the entire systemic circulatory system may become so great that fluid leaks continually into the tissue

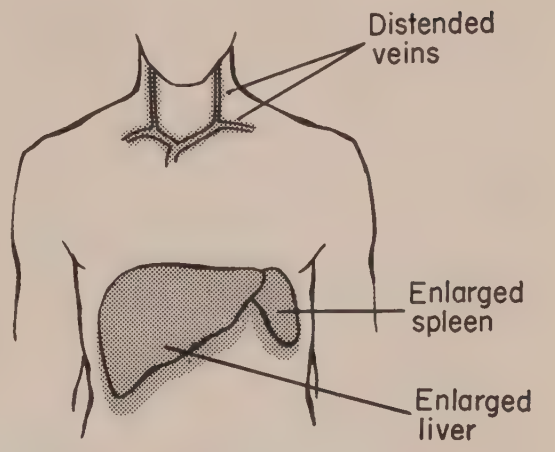


Figure 147. Engorgement of the peripheral venous system and blood reservoirs in cardiac failure.

spaces, resulting in extreme *generalized edema.* One of the conditions frequently encountered in severe heart disease is swelling from head to toe as the result of excessive interstitial fluid, a condition called *dropsy.*

The edema in cardiac failure is caused not only by high pressure in the capillaries but also by two other factors: First, decreased cardiac output decreases the amount of urine formed by the kidneys, causing excessive quantities of water and electrolytes to remain in the body, enhancing the volume of extracellular fluid. Second, diminished blood flow through the body causes the adrenal cortex to secrete large quantities of aldosterone which also makes the kidneys retain water and salt, thus further increasing the total extracellular fluid volume.

One of the most essential features in the treatment of cardiac failure is the control of water and salt intake. Also, very powerful drugs called *diuretics,* which increase the output of urine by the kidneys, are frequently administered.

CIRCULATORY SHOCK

Circulatory shock is the condition resulting when the cardiac output becomes so reduced that the tissues everywhere or almost everywhere in the body fail to receive an adequate blood supply. The tissues suffer from inadequate nutrition and inadequate removal of cellular excretory products because of the reduced blood flow.

Any condition that decreases the cardiac output to a very low value can cause circulatory shock. Therefore, shock can result from weakness of the heart itself or from diminished venous

return, for which reason it can be classified into two main types, (1) *cardiac shock*, and (2) *low venous return shock*.

Cardiac Shock

Cardiac shock is caused by decreased effectiveness of the heart as a pump. This type of shock occurs most frequently immediately after a severe heart attack because the heart's ability to pump blood often decreases many-fold in only a few minutes. The person frequently dies because of diminished blood supply to his tissues before the heart can begin to recover. In essence, then, cardiac shock is a severe degree of low cardiac output failure.

Shock Caused by Diminished Venous Return

Any of the factors that decrease the tendency for blood to return to the heart can cause shock. These are (1) decreased blood volume, which causes *hypovolemic shock*, (2) increased size of the vascular bed so that even normal amounts of blood will not fill the vessels adequately—this is called *venous pooling shock*, or (3) obstruction of blood vessels, particularly veins.

Hypovolemic shock. The blood volume can be decreased as a result of hemorrhage, loss of plasma through exuding wounds or burns, loss of plasma into severely crushed tissues, or dehydration caused by extreme sweating, lack of water to drink, or excessive loss of fluids through the gut or kidneys. In any of these conditions, the diminished blood volume decreases the mean systemic pressure so greatly that inadequate quantities of blood return to the heart, and the diminished venous return causes shock.

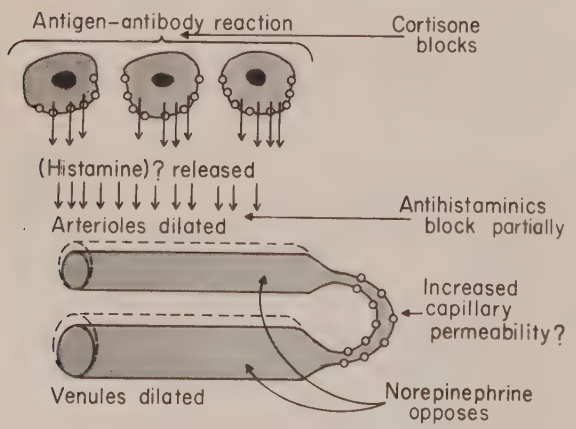


Figure 148. Mechanism of anaphylactic shock.

Venous pooling shock. If the blood vessels lose their vasomotor tone, their diameters may increase so greatly that the blood collects, or "pools," in the highly distensible veins. As a result, the pressures in the systemic circulation fall so low that little *pressure gradient for venous return* then exists to make the blood flow toward the heart, and venous return becomes so slight that circulatory shock ensues.

A special type of venous pooling shock is *neurogenic circulatory shock* caused by sudden cessation of sympathetic impulses from the central nervous system to the peripheral vascular system. The result is loss of normal vasomotor tone, diminished pressures everywhere in the systemic circulation, and consequently diminished venous return. Fainting is an example of this type of shock.

Extreme allergic reactions can also cause venous pooling shock. This type of shock is known as *anaphylactic shock*, and its probable mechanism is that illustrated in Figure 148, which may be explained as follows: When a person becomes immune to a disease, his body manufactures *antibodies;* these are special proteins that destroy bacteria or toxins. Occasionally the reaction of the antibodies with substances entering the circulation damages the body's own tissues, at the same time causing a substance called *histamine* to be released into the body fluids. The histamine then circulates to all the peripheral blood vessels and promotes vasodilatation, venous pooling, and diminished venous return, culminating immediately in a state of anaphylactic shock. Indeed, venous pooling can sometimes result so rapidly and to such an extreme extent during anaphylaxis that the person may die before therapy can be started.

Shock caused by obstruction to venous return. Very rarely shock is caused by obstruction of blood vessels, which can prevent adequate return of blood to the heart from the systemic circulation. This occurs particularly when large veins are blocked, as when growths press on the inferior vena cava in the region of the liver or in the abdomen.

Stages of Shock

Compensatory stage. If the abnormality causing shock is mild, the regulatory mechanisms of the circulatory system can compensate for it, and no damage will result. Figure 149 illustrates this ability of the circulatory system to compensate for hemorrhage. At the bottom of the figure, the small circles represent the changing size of a representative blood vessel. As more and more

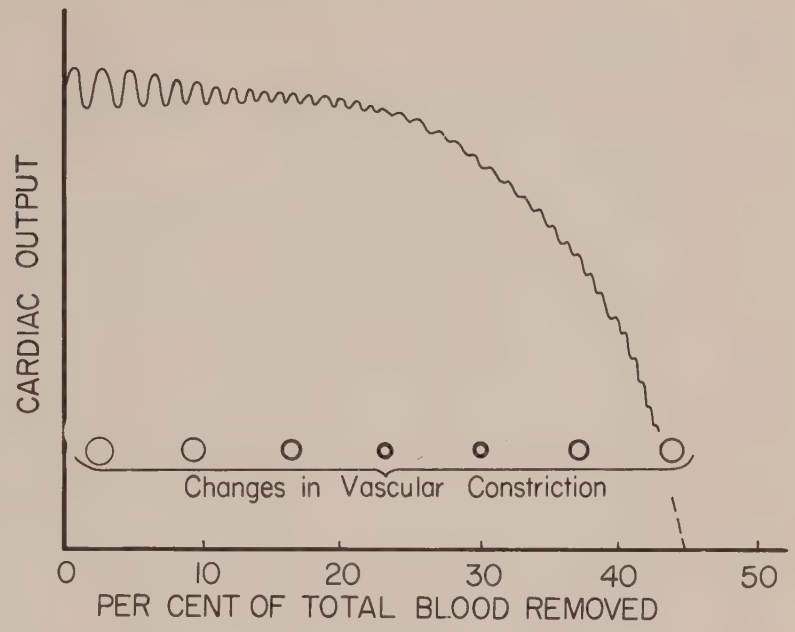


Figure 149. Effect of blood loss on cardiac output, showing compensatory vascular constriction.

blood is removed from the circulation, the blood loss is compensated by constriction of the veins and venous reservoirs. Consequently, even though blood is lost, the return of blood to the heart continues as usual, and the cardiac output remains essentially normal until 15 to 20 per cent of the total blood volume (about 900 ml.) is removed. However, beyond this point, the blood loss can no longer be compensated, and cardiac output begins to fall. This early stage of shock is called the *compensatory stage.*

Fortunately, the arterioles supplying the heart do not constrict during the compensatory stage of shock, and those supplying the brain constrict very little, so that blood continues to flow to these essential organs in normal or almost normal quantities. Consequently, during this stage of shock, the person is not in imminent danger of dying, and unless the cause of the shock is intensified, the condition will be corrected and the person will return to normal within a few hours or certainly within a day or two.

Progressive stage. If the degree of shock becomes very severe, regardless of the initial cause of the shock, *the shock itself promotes more shock.* Some of the reasons for this are shown in Figure 150. First, if the degree of shock is so great that the heart fails to pump enough blood to supply its own coronary vessels, the heart itself becomes progressively weakened. This further diminishes the cardiac output, which weakens the heart even more and diminishes the cardiac output again, leading to a vicious cycle that eventually kills the person.

Second, very poor blood flow to the brain causes damage to the vasomotor and respiratory centers. Failure of the vasomotor center allows venous pooling, resulting in even more extensive shock, and respiratory failure causes diminished oxygenation of the blood and, therefore, malnu-

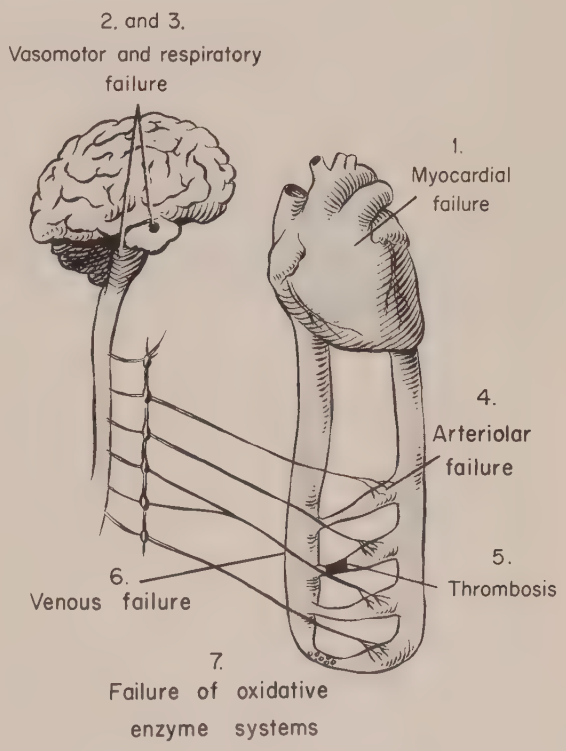


Figure 150. Factors that cause progression of shock and finally cause it to become irreversible.

trition of the tissues, which also makes the shock worse.

Peripheral vascular failure is a third cause of a vicious cycle. Ischemia of the blood vessels can probably make the vascular musculature so weak that the vessels dilate, resulting in vascular pooling of blood, diminished venous return, increased shock, increased ischemia, and so forth, the vicious cycle repeating itself again and again until the vessels are fully dilated or the patient is dead.

Fourth, recent experiments have shown that greatly diminished blood flow causes minute blood clots to develop in the small vessels. As a consequence of the plugged vessels, venous return decreases, and the progressively more sluggish flow of blood through the peripheral vessels causes more and more clotting, creating still another vicious cycle.

Fifth, it is believed that greatly prolonged shock might damage the tissue enzymes, which causes all the circulatory functions to diminish further.

Thus, because of several vicious cycles that can develop, once shock has reached a certain degree of severity, progressive deterioration of the circulatory system causes the shock to become more and more severe until death ensues.

IRREVERSIBLE SHOCK. In the early stages of progressive shock, the person's life can usually be saved by appropriate and rapid treatment. For instance, if the shock is initiated by hemorrhage, rapid transfusion of blood can restore normal circulatory dynamics unless the damage to the circulation is already very great. However, if the shock has progressed for a long period of time, such extreme damage may have occurred that no amount of treatment can be successful in restoring the cardiac output enough to sustain life. When this state has been reached, the patient is said to be in *irreversible shock*. The heart, for instance, may have become damaged beyond repair by this time so that it is incapable of pumping more than 1 to 1.5 liters per minute regardless of what therapy is instituted. Or so many peripheral vessels may have become plugged with small clots that blood can no longer flow rapidly enough to repair the damage. In these instances, despite any amount of treatment, the vicious cycles of the progressive stage of shock continue on and on until death of the person.

Treatment of Shock

The treatment of shock depends on the cause. In the case of cardiac shock, the best therapy is to increase the pumping effectiveness of the heart. Usually this is not easily accomplished, but bed rest and the drug digitalis, which often strengthens a weakened heart muscle, are of value. Also, one can sometimes help a person with cardiac shock by judiciously transfusing blood. This will increase the pressures in the systemic circulation and thereby force an increased quantity of blood into the heart. Even a weakened heart can usually pump increased blood when its chambers are stretched a little.

If the shock results from hypovolemia, the blood volume can be increased by administering blood, plasma, or sometimes even isotonic salt solution. Return of the blood volume to normal increases the pressures throughout the systemic vessels, which brings the venous return back to normal and thereby alleviates the shock.

In both neurogenic and anaphylactic shock, the main difficulty is venous pooling. Administration of drugs that act like the sympathetic nervous system in constricting the blood vessels—such as norepinephrine—compresses the blood in the vessels and forces it along the veins toward the heart, bringing the patient out of shock.

REFERENCES

Altschule, M. D.: *Physiology in Diseases of the Heart and Lungs.* Cambridge, Harvard University Press, 1954.

Brecher, G. A.: *Venous Return.* New York, Grune & Stratton, Inc., 1956.

Friedberg, C. K.: *Diseases of the Heart.* 2nd Ed. Philadelphia, W. B. Saunders Company, 1956.

Gregersen, M. I., and Rawson, R. A.: Blood volume. *Physiol. Rev.* 39:307, 1959.

Guyton, A. C.: *Circulatory Physiology: Cardiac Output and Its Regulation.* Philadelphia, W. B. Saunders Company, 1963.

Guyton, A. C.: Determination of cardiac output by equating venous return curves with cardiac response curves. *Physiol. Rev.* 35:123, 1955.

Guyton, A. C., and Adkins, L.: Quantitative aspects of the collapse factor in relation to venous return. *Am. J. Physiol.* 177:523, 1954.

Guyton, A. C., Armstrong, G. G., and Chipley, P. L.: Pressure-volume curves of the arterial and venous systems in live dogs. *Am. J. Physiol.* 184:253, 1956.

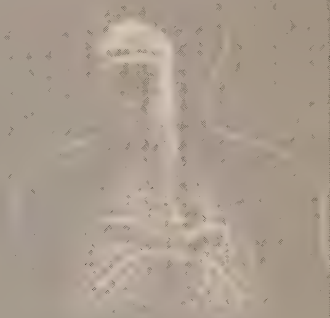
Guyton, A. C., Batson, H. M., and Smith, C. M.: Adjustments of the circulatory system following very rapid transfusion or hemorrhage. *Am. J. Physiol.* 164:351, 1951.

Hamilton, W. F.: Heart output. In Glasser, O., editor: *Medical Physics.* Vol. 1. Chicago, Year Book Publishers, Inc., 1944.

Reeve, E. B., Allen, T. H., and Roberts, J. E.: Blood volume regulation. *Ann. Rev. Physiol.* 22:349, 1960.

Section Five

RESPIRATION



MECHANICS OF RESPIRATION AND TRANSPORT OF OXYGEN AND CARBON DIOXIDE

17

The respiratory system, shown in Figure 151, is composed of the lungs, trachea, glottis, pharynx, and nose. This system supplies oxygen to the blood and removes carbon dioxide. To the upper right in the figure is shown a pulmonary capillary in close contact with an alveolus of the lung. It is through the membrane separating the alveoli and the capillaries that oxygen leaves the air to enter the blood and carbon dioxide returns to the air. The object of breathing is to move air continually into and out of the alveoli, thereby maintaining a fresh supply of alveolar oxygen and removing carbon dioxide as it accumulates.

FUNCTIONS OF THE RESPIRATORY PASSAGEWAYS

Functions of the Nose

The inside surface of the nose is extensive. The nasal cavity is divided by a central septum, and several projections called *turbinates* extend into each cavity from the lateral side. Air passing through the nose, on coming in contact with all of the nasal surfaces, becomes warmed and humidified. The turbinates also cause many eddies in

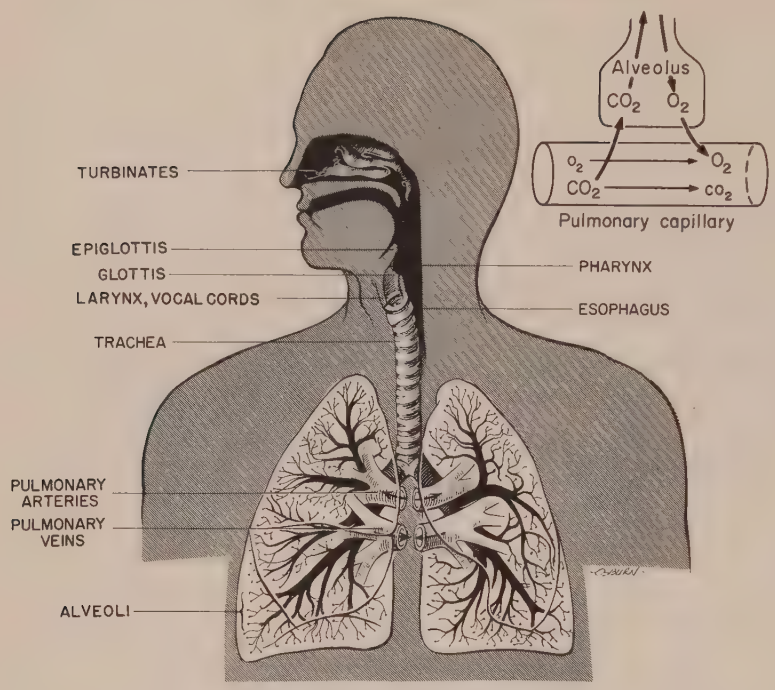


Figure 151. The respiratory system, showing the respiratory passages and function of the alveolus to oxygenate the blood and to remove carbon dioxide.

the flowing air, forcing it to rebound in many different directions before finally completing its passage through the nose. This causes small dust or other suspended particles in the air to precipitate against the nasal surfaces by the following mechanism: When air containing a foreign particle travels toward a surface and then suddenly changes its direction of movement, the momentum of the particle makes it continue to travel in the original direction, while the air, which has little momentum because of its low mass, takes off in the new direction. The particles passing through the nose impinge on a turbinate or on another surface of the nasal passageways and are entrapped in the layer of mucus covering the surface. This method of removing foreign particles from the air is so efficient that very rarely do particles of greater size than 4 to 6 microns pass through the nose into the lower respiratory passageways.

Functions of the Larynx

The pharynx separates into the trachea and esophagus immediately above the larynx. Here food is separated from air, and the air passes into the *trachea* while the food passes into the *esophagus.*

The separation of food and air is controlled by local nerve reflexes. Whenever food touches the surface of the *pharynx,* the *epiglottis* automatically closes over the opening of the trachea, allowing the food to slide on into the esophagus. When air passes into the pharynx, the epiglottis remains open, and the air takes the course of least resistance into the trachea and lungs.

Function of the vocal cords. The vocal cords are the portion of the larynx that make sound. The *vocal cords* are two small vanes located on either side of the air passageway, as shown in Figure 152. Contraction of muscles in the larynx can bring

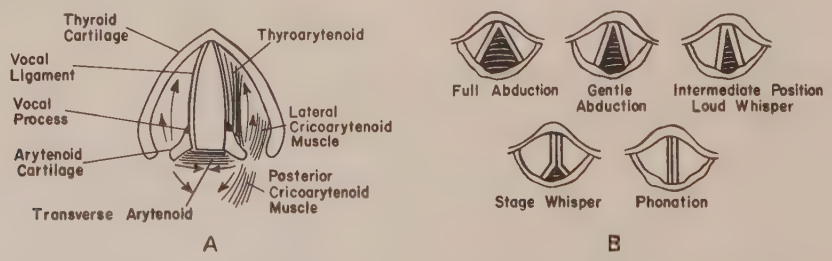


Figure 152. Laryngeal function in phonation. (Modified from Greene: *The Voice and Its Disorders.* Pitman Medical Publishing Co.)

these vanes close to each other or can spread them apart. They can also be stretched or relaxed. When the cords are approximated and air is forced between them, they vibrate to generate sound, and the different *pitches* of sound are controlled by the degree to which the cords are stretched. The formation of words or other complicated sounds is a function of the mouth as well as the larynx because the *quality* of a sound depends upon the momentary position of the lips, cheeks, teeth, tongue, and palate.

For speech or other sounds to be emitted, the respiration, the vocal cords, and the mouth must all be controlled at the same time. This is effected by a special brain center, called *Broca's area,* located in the left frontal lobe of the brain. The organization and function of this center will be discussed in Chapter 24.

The Cough and Sneeze Reflexes

A means for keeping the respiratory passages clean is to force air very rapidly outward by either coughing or sneezing. The cough reflex is initiated by any irritant touching a surface of the glottis, the trachea, or a bronchus. Sensory impulses are transmitted to the medulla of the brain, and, in turn, appropriate impulses are transmitted back to the respiratory system and larynx to cause the cough. The respiratory muscles first contract very strongly, building up high pressure in the lungs, while at the same time the vocal cords remain clamped tightly closed. Then suddenly the vocal cords open, allowing the pressurized air to flow with a blast out of the lungs. Sometimes the air flows as rapidly as 70 miles per hour. In this way unwanted foreign matter such as particles, mucus, or other substances are expelled from the respiratory passageways.

The sneeze reflex is very similar to the cough reflex except that it is initiated by irritants in the nose. Impulses pass from the nose to the medulla and then back to the respiratory system. A sudden forceful expiration blows air outward while the soft palate varies its position to allow rapid flow of air successively through the nose and mouth. In this way the sneeze is capable of clearing the nasal passageways in the same manner that the cough reflex clears many of the lower passageways.

FLOW OF AIR INTO AND OUT OF THE LUNGS

The lungs are enclosed in the *thoracic cage,* which is comprised of the *sternum* in front, the *spinal column* in back, and the *ribs* encircling the chest. The act of breathing is performed by enlarging and contracting the thoracic cage. The cavity formed by the thoracic cage is called the *pleural cavity,* and the lungs normally fill this cavity entirely. The lungs are covered with a lubricated membrane called the *visceral pleura,* and the inside of the pleural cavity is lined with a similar membrane called the *parietal pleura.* The

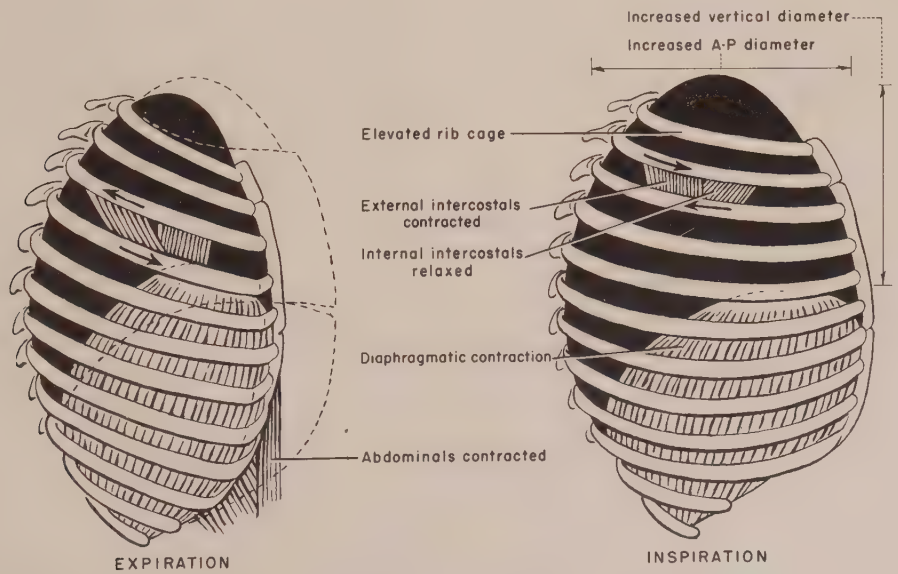


Figure 153. The major muscles of expiration and inspiration, and changes in the thoracic cage during expiration and inspiration.

lungs slide freely inside the pleural cavity, so that any time the cavity enlarges, the lungs must also enlarge. In other words, any change in the volume of the thoracic cage is immediately reflected by a similar change in the volume of the lungs.

The muscles of respiration. Figure 153 illustrates several of the muscles of respiration, and also the changes in shape of the thoracic cage during inspiration and expiration. The major muscles of inspiration are the *diaphragm,* the *external intercostals,* and a number of small muscles in the neck which pull upward on the front of the thoracic cage. The inspiratory muscles cause the pleural cavity to enlarge in two ways. First, downward movement of the diaphragm pulls the bottom of the pleural cavity downward, thus elongating it, which is shown to the right in Figure 153. Second, the external intercostals and neck muscles lift the front of the thoracic cage, causing the ribs to angulate more directly forward than previously, increasing the thickness of the cage, which is also shown in the right half of the figure.

The major muscles of expiration are the *abdominals* and, to a lesser extent, the *internal intercostals.* The abdominal muscles cause expiration in two ways. First, they *pull downward on the chest cage* thereby decreasing the thoracic thickness. Second, they *force the abdominal contents upward* against the diaphragm, decreasing the longitudinal dimension of the pleural cavity. The internal intercostals help in the process of expiration to a slight extent by pulling the ribs downward. When they are in this downward position, the thickness of the chest is considerably decreased, as can be seen in the left half of Figure 153.

Alveolar and Intrapleural Pressures

During inspiration the thoracic cage enlarges, which also enlarges the lungs. One will recall from the basic laws of physics that when a volume of gas is suddenly enlarged, its pressure falls. Thus, during inspiration the enlargement of the thoracic cage decreases the pressure in the alveoli to about −3 mm. Hg, and it is this negative pressure that pulls air through the respiratory passageways into the alveoli.

During expiration exactly the opposite effects occur; compression of the thoracic cage around the lungs increases the alveolar pressure to approximately +3 mm. Hg, which obviously pushes air out of the alveoli to the atmosphere.

If a person breathes with maximal effort but with his nose and mouth closed so that air cannot flow into or out of his lungs, the alveolar pressure can be decreased to as low as −80 mm. Hg or increased to as high as +100 mm. Hg. Thus, the maximum power of the muscles of respiration is far greater than that needed for normal quiet respiration. This provides a tremendous reserve respiratory ability which can be called upon in times of need for maximal respiratory activity, such as during heavy exercise.

The space between the lungs and the thoracic cage is called the *intrapleural space,* and the pressure in this space is called the *intrapleural pressure.* This pressure is always a few millimeters of mercury less than that in the alveoli. Figure 154 illustrates that during inspiration the intra-alveolar pressure is about −3 mm. Hg while the intrapleural pressure is about −8 mm. Hg. During normal expiration, the intra-alveolar pressure rises to +3 mm. Hg while the intrapleural pressure rises to −2 mm. Hg. It will be noted, therefore, that the intrapleural pressure remains about 5 mm. Hg less than the intra-aveolar pressure all the time. The reason for this is two-fold: First, the surface tension of the fluid lining the inside of the alveoli is always pulling the alveolar surfaces into the smallest possible bubble size. Second, elastic fibers spread in all directions throughout the tissues of the lungs, and these also tend to contract the lungs at all times. The effect of both of these is to pull the lungs away from the thoracic cage, creating an average negative pressure in the intrapleural space of about −5 mm. Hg.

PNEUMOTHORAX. When an opening is made in the chest wall the elastic forces in the lungs cause them to collapse immediately, sucking air through the opening into the chest cavity. Then, air flows in and out of the hole in the chest when the person tries to breathe, and his lungs do not expand and contract. Thus, a wound of the chest can kill a person by suffocation; yet the condition can be treated very easily simply by sucking air out of the pleural cavity and plugging the hole.

"SURFACE ACTIVE AGENT" IN THE ALVEOLI. If the alveoli were lined with water instead of the normal alveolar fluid, the surface tension would be so great that it would cause the alveoli to remain collapsed almost all the time. However, a substance called a *surface active agent* is secreted into the alveoli through the alveolar membrane. This substance is actually a detergent that greatly decreases the surface tension of the fluid lining the alveoli.

Artificial Respiration

When the respiratory muscles fail, a person can be kept alive only by some means of artificial respiration. The simplest method is to force air

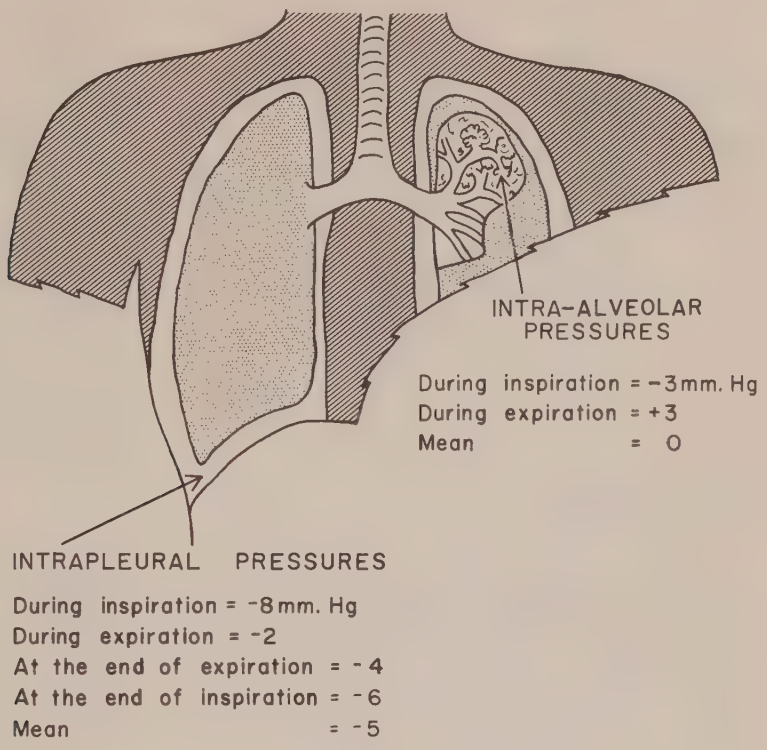


Figure 154. Alveolar and intrapleural pressures during normal breathing.

into and out of his mouth or nose. Several devices have been made for this purpose, one of which is called the *resuscitator.* To use this apparatus a mask is placed over the mouth and nose and intermittent blasts of air fill the lungs. Between blasts, air is pulled back out of the lungs into the atmosphere.

For prolonged artificial respiration the person is usually placed inside an artificial respirator of the type shown in Figure 155. This is a large tank provided with some means for repetitively increasing and decreasing the pressure and for controlling the extent of the rise and fall in pressure. In the respirator of Figure 155 the air pressure is alternated by inward and outward movement of a large leather diaphragm, and the positive and negative pressures are controlled by the relief valves shown on top of the tank. Because respiration normally is accomplished by the inspiratory muscles rather than by the expiratory muscles, the pressure inside the tank is usually adjusted to give more of a vacuum cycle than a pressure cycle. A vacuum of about −10 mm. Hg causes approximately normal inspiration by pulling outward on the chest cage and abdominal wall, thereby sucking air into the lungs. Then, the respirator changes from the vacuum to the pressure cycle, applying 2 to 3 mm. Hg of positive pressure. This causes expiration by pushing against the abdomen and chest.

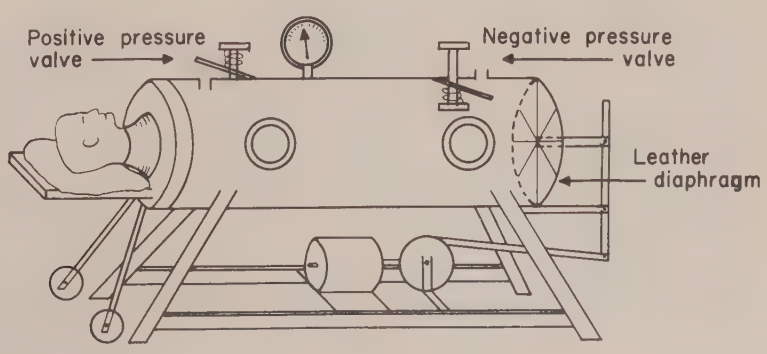


Figure 155. The artificial respirator.

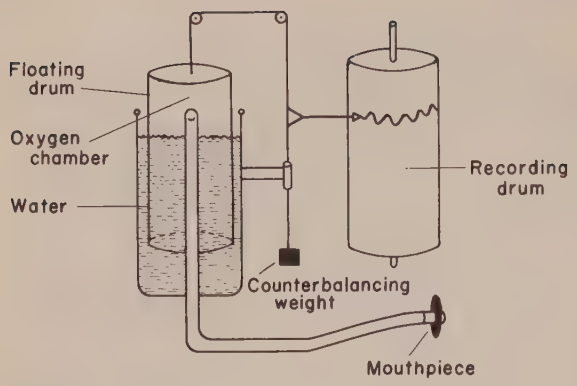


Figure 156. The spirometer.

Spirometry and the Divisions of the Respiratory Air

Figure 156 represents a *spirometer,* an apparatus that can be used to record the flow of air into and out of the lungs. It is composed of an inverted floating drum with a tube extending from the air space in the drum to the mouth of the subject. As he breathes in and out, the drum moves up and down, and a counterweight balancing the drum also rides up and down, recording on a moving paper the changing volume of the chamber. Figure 157 illustrates a typical spirometric recording.

The tidal volume, respiratory rate, and the minute respiratory volume. The air that passes into and out of the lungs with each respiration is called the *tidal air,* and the volume of this air in each breath is called the *tidal volume.* The normal tidal volume is about 500 ml., and the normal rate of respiration for an adult is usually about 12 times per minute. Therefore, a total of about 6 liters of air normally passes into and out of the respiratory passageways each minute. This amount is called the *minute respiratory volume.*

The inspiratory capacity. In Figure 157, after three normal breaths, the subject breathes inward as deeply as possible. The amount of air that he can pull into his lungs beyond that already in his lungs at the beginning of the breath is called the *inspiratory capacity.* This quantity is approximately 3000 ml. in the normal person.

Expiratory reserve volume. After several more normal respirations, the subject expires as much as possible (see Figure 157). The amount of air that he is capable of expiring beyond that which he normally expires is called the *expiratory reserve volume.* This is usually about 1100 ml.

RESIDUAL VOLUME AND THE FUNCTIONAL RESIDUAL CAPACITY. In addition to the expiratory reserve volume, there is air in the lungs that cannot be expired even by the most forceful exhalation. The volume of this air is about 1200 ml., and it is called the *residual volume.* The sum of the expiratory reserve volume plus the residual volume is called the *functional residual capacity;* this is the amount of air remaining in the respiratory system at the end of a normal expiration. It is this air that allows oxygen and carbon dioxide transfer into and out of the blood to continue even between periods of inspiration.

Vital capacity. Finally, Figure 157 shows the spirogram of a maximal inspiratory effort followed immediately by a maximal expiratory effort. The total change in pulmonary volume between these two extremes is called the *vital capacity.* The vital capacity of the normal person is approximately 4500 ml. A well trained male athlete may have a

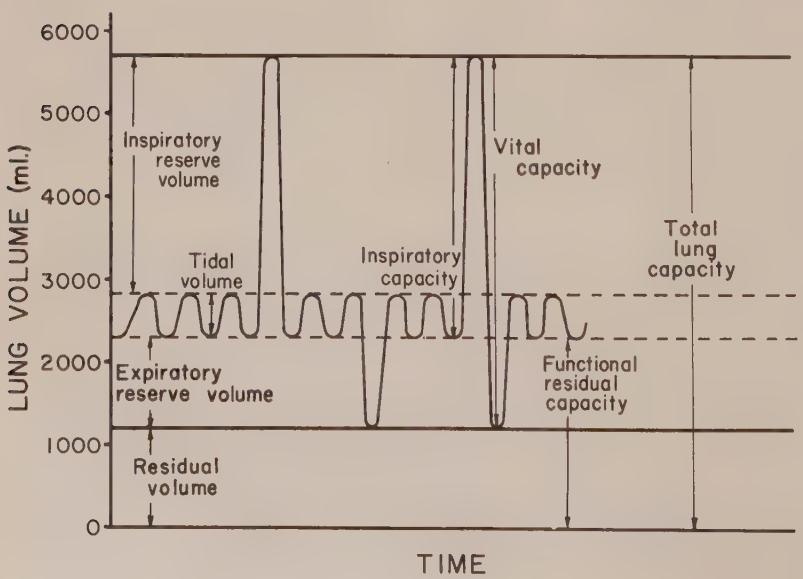


Figure 157. A spirogram, showing the divisions of the respiratory air.

vital capacity as great as 6500 ml., while an asthenic female frequently has a vital capacity no greater than 3000 ml.

The vital capacity is a measure of the person's overall ability to inspire and expire air, and it is determined by two factors: (1) the strength of the respiratory muscles, and (2) the resistance of the thoracic cage and lungs to expansion and contraction. Disease processes that either weaken the muscles—such as poliomyelitis—or decrease the expansibility of the lungs—such as tuberculosis—can decrease the vital capacity. For this reason vital capacity measurements are an invaluable tool in assessing the functional ability of the lungs.

Dead Space

Much of the air pulled into the respiratory passages with each breath is expired again before it ever reaches the alveoli. This air is useless from the point of view of oxygenating the blood. Consequently the respiratory passageways are called *dead space.* The total volume of this space is normally about 150 ml., which means that during inspiration of a normal tidal volume of 500 ml., only 350 ml. of new air actually enters the alveoli.

Alveolar ventilation. The most important measure of the effectiveness of a person's respiration is his *alveolar ventilation,* which is the total quantity of new air that enters his alveoli each minute. With only 350 ml. of air reaching the alveoli with each breath and with a normal respiratory rate of 12 times per minute, the alveolar ventilation is approximately 4200 ml. per minute. During maximal respiratory effort this can be increased to as high as 100 liters per minute, and at the opposite extreme a person can remain alive at least for a few hours with an alveolar ventilation as low as 1200 ml. per minute.

Exchange of Alveolar Air with Atmospheric Air

At the end of each expiration approximately 2300 ml. of air still remains in the lungs. This is called the *functional residual capacity,* as was explained previously. With each breath, about 350 ml. of new air is brought into this air, and with each expiration this same amount of air is removed. It is obvious, then, that each breath does not change the total content of the alveolar air, but instead *changes only about one seventh of the air.* This effect is illustrated in Figure 158, which shows an alveolus containing a quantity of some foreign gas at the beginning of the series of breaths. After the first breath a small portion of this gas has been removed, after the second breath a little more, after the third a little more, and so forth. Note that even after the sixteenth breath a small portion of the foreign gas is still in the alveolus. At a normal alveolar ventilation of 4200 ml. per minute, approximately one-half the alveolar gases are replaced by new air every 23 seconds. This slow turnover of air in the alveoli makes it possible to maintain alveolar gaseous concentrations that do not rise and fall greatly with each individual breath.

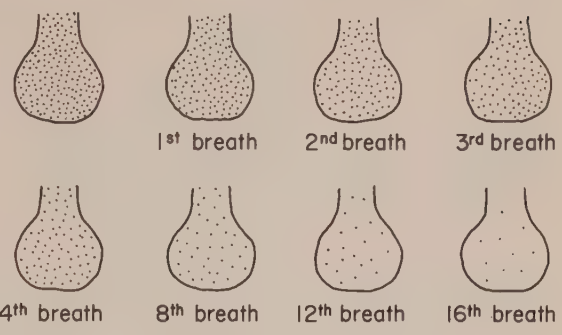


Figure 158. Expiration of a foreign gas from the alveoli with successive breaths.

ALVEOLAR AIR

The air in the alveoli is separated from the blood in the capillaries by the pulmonary membrane which has a thickness of only 0.2 to 0.4 micron. Therefore, oxygen can diffuse readily from the alveolar air into the blood, and carbon dioxide can diffuse from the blood into the alveolar air. One of the principal factors that determines the rate at which a gas will diffuse through this membrane is the *difference between the partial pressures* of the gas on the two sides of the membrane. Because many students may not be familiar with the concept of partial pressure, it is essential that this be explained at this point.

Partial Pressures

Figure 159 shows four chambers of the same size. Chamber B contains two separate compartments, the upper one storing 79 volumes of nitrogen and the lower one 21 volumes of oxygen. The pressure in each of these compartments is normal atmospheric pressure, 760 mm. Hg. The nitrogen is moved to chamber A which holds 100 volumes. In other words, the nitrogen expands from 79 to 100 volumes. In this process the pressure falls to

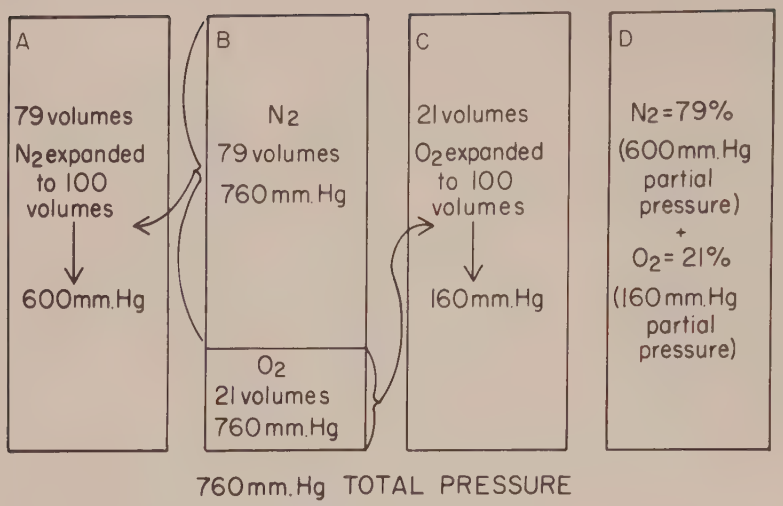


Figure 159. Partial pressures of oxygen and nitrogen in different chambers (explained in the text).

79/100 its original value, or from 760 mm. to 600 mm. Similarly, the oxygen expands when moved to chamber C, and its pressure falls from 760 to 160. Finally, in chamber D, the nitrogen and oxygen in chambers A and C are mixed together. Now it is found that the total pressure in this chamber is once again 760 mm. Hg. Reasoning from the pressure and volume changes in these chambers, it is evident that 600 mm. of the 760 mm. Hg pressure in the fourth chamber is caused by nitrogen and 160 mm. is caused by oxygen. Therefore, the *partial pressure of nitrogen* (P_{N_2}) in this mixture is said to be 600 mm. Hg, while the partial pressure of oxygen (P_{O_2}) is 160 mm. Hg.

The partial pressure of a gas is a measure of the total force that it exerts against the walls surrounding it. Consequently, the penetrating power of a gas is directly proportional to its partial pressure. The greater the partial pressure of a gas in the alveolus, the greater is its tendency to pass through the pulmonary membrane into the blood.

Table 3 shows the relative pressures of nitrogen, oxygen, carbon dioxide, and water vapor in atmospheric and alveolar airs. This table shows that normal atmospheric air contains about four-fifths nitrogen and one-fifth oxygen, with almost negligible quantities of carbon dioxide and water vapor. On the other hand, alveolar air contains considerable quantities of both carbon dioxide and water vapor, while the oxygen content is considerably less than that of atmospheric air. These differences can be explained as follows:

Table 3. *Partial Pressures of Respiratory Gases in the Atmosphere and in the Alveoli*

Gas	*Atmospheric Air*	*Alveolar Air*
N_2	597.0 (78.62%)	569.0 (74.9%)
O_2	159.0 (20.84%)	104.0 (13.6%)
CO_2	0.15 (0.04%)	40.0 (5.3%)
H_2O	3.85 (0.5%)	47.0 (6.2%)

Humidification of air as it enters the respiratory passages. When air is inspired it is humidified immediately by the moisture on the linings of the respiratory passages and alveoli. At normal body temperature the partial pressure of water vapor in the lungs is 47 mm. Hg, and as long as the body temperature remains constant, this partial pressure also remains constant.

The mixing of water vapor with the incoming atmospheric air dilutes the air so that the pressure of the other gases becomes slightly decreased. This explains why the nitrogen partial pressure in the alveoli is slightly less than its partial pressure in atmospheric air.

Oxygen and carbon dioxide partial pressures in the alveoli. Alveolar air continually loses oxygen to the blood, and this oxygen is replaced by carbon dioxide diffusing out of the blood into the alveoli. This explains why the oxygen pressure in alveolar air is much less than in atmospheric air, and it also explains why the carbon dioxide pressure is much greater in the alveolar air than in the atmospheric air. The normal partial pressure of oxygen in the alveoli is approximately 104 mm. Hg, and that of carbon dioxide is 40 mm. Hg. However, these values change greatly from time to time depending on the rate of alveolar ventilation and the rate of oxygen and carbon dioxide transfer into and out of the blood. For instance, a high alveolar ventilation provides increased amounts of oxygen to the alveoli and increases the alveolar oxygen pressure. Also, a high alveolar ventilation removes carbon dioxide from the alveoli more rapidly than usual, thereby decreasing the carbon dioxide pressure.

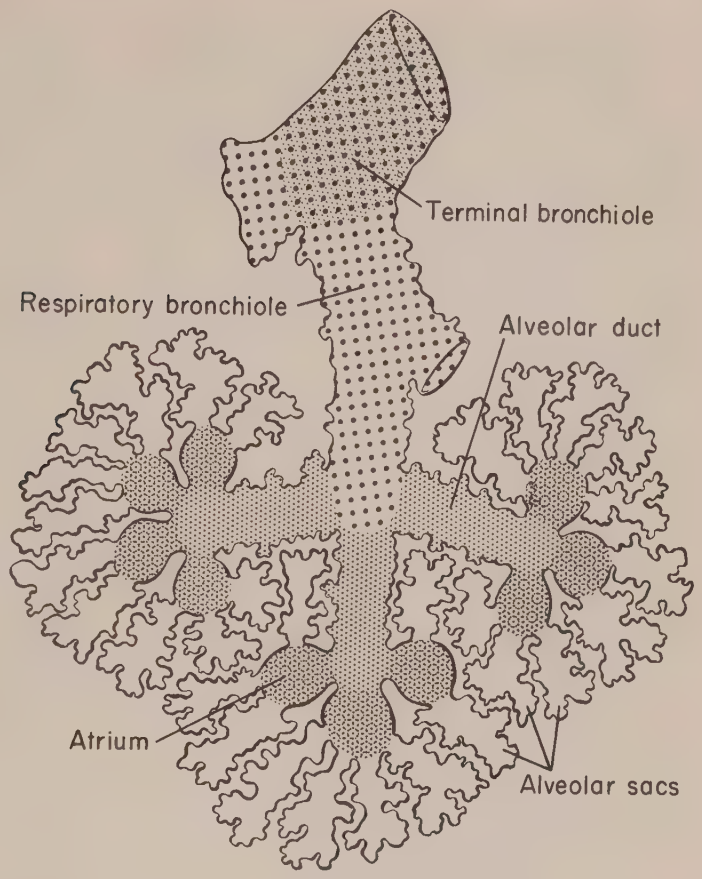


Figure 160. The respiratory unit, illustrating the anatomy of the respiratory surfaces. (From Miller: *The Lung*. Charles C Thomas.)

TRANSPORT OF GASES THROUGH THE PULMONARY MEMBRANE

The Pulmonary Membrane

The pulmonary membrane is comprised of all the pulmonary surfaces that are thin enough to allow gases to diffuse into the pulmonary blood. These include, as illustrated in Figure 160, the membranes of the respiratory bronchioles, the alveolar ducts, the atria, and the alveolar sacs.

The total area of the pulmonary membrane is approximately 60 sq. m., which is about equal to the floor area of a moderate sized classroom. Only about 60 ml. of blood is in the capillaries at any one time. If one will think for a moment of this small amount of blood spread out evenly over the entire floor of a classroom, he can quite readily understand that very large quantities of gas can enter and leave this blood in only a fraction of a second.

The thickness of the pulmonary membrane is only 1 to 4 microns, which is several times less than the thickness of a red blood cell. This, also, is one of the features of the pulmonary membrane that allows extremely rapid diffusion between the alveolar air and the blood.

Solution and Diffusion of Gases in Water

In order to understand the transport of gases through the pulmonary membrane, one must first be familiar with the physical principles of gaseous solution and diffusion in water. Figure 161 shows a chamber containing water at the bottom and normal air at the top. Molecules of nitrogen and oxygen are continually bouncing against the sur-

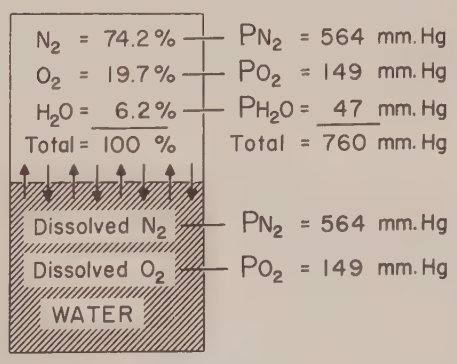


Figure 161. A sample of air exposed to water, showing equilibration between gases in the gaseous phase and in the dissolved state.

face of the water, and some of the molecules enter the water to become dissolved. The dissolved molecules bounce among the water molecules in all directions, some of them eventually reaching the surface again to bounce back into the gaseous space. After the air has remained in contact with the water for a long time, the number of molecules passing out of the solution equals exactly the number of molecules passing into the solution. When this state has been reached, the gases in the gaseous phase are said to be in equilibrium with the dissolved gases.

In the equilibrium state, the partial pressure tending to force nitrogen molecules into the water is 564 mm. Hg, but the pressure of nitrogen molecules pushing outward from the water is also 564 mm. Hg. If the nitrogen pressure becomes greater in the gaseous phase than in the dissolved phase, more nitrogen molecules will move into the solution than out. If the pressure becomes less in the gas than in the solution, more nitrogen molecules will leave the solution than will enter it. In other words, *the pressure of a gas is the force with which it attempts to move from its present surroundings.* Whenever the pressure is greater at one point than at another point, whether this be in a solution or in a gaseous mixture, more molecules will move toward the low pressure area than toward the high pressure area.

Diffusion of gases through tissues and fluids. Figure 162 shows a chamber containing a solution. At point A is a high concentration of the dissolved molecules, while at point B is a low concentration. The molecules are bouncing at random in all directions. Obviously, because of the higher concentration at point A than at point B, it is easier for large numbers of molecules to move from A to B than in the opposite direction, which is indicated by the lengths of the arrows, and eventually the numbers of molecules at both ends of the chamber will be approximately equal. Movement of molecules in this manner from areas of high concentration toward areas of low concentration is called *diffusion.*

The rate of diffusion of molecules in the body fluids and in the body tissues is determined by the pressure differences between the different points. Far more molecules are at point A in Figure 162 than at point B. Consequently, the pressure of the gas at point A is also far greater than at point B. The difference between the two pressures is called the *pressure gradient,* and the rate of diffusion between the two points is directly proportional to the pressure gradient.

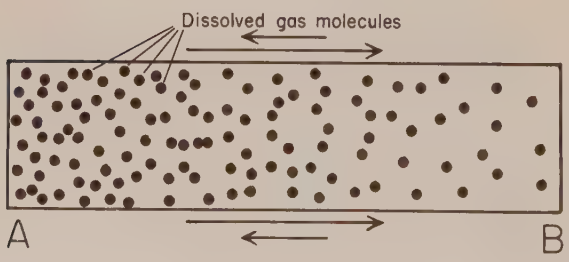


Figure 162. Diffusion of gas molecules through liquids.

Diffusion of gases through the pulmonary membrane. The factors that determine the rate of gas diffusion through the pulmonary membrane are the following:

1. The greater the pressure gradient between one side of the membrane and the other, the greater will be the rate of gaseous flow. If the pressure of a gas in the alveolus is 100 mm. Hg while that in the blood is 99 mm. Hg, the pressure gradient will be only 1 mm., and the rate of gaseous flow will be very slight. If the pressure in the blood suddenly falls to 0 mm. Hg, the pressure difference rises to 100 mm., and the gaseous flow increases to 100 times its former rate.

2. The greater the area of the pulmonary membrane, the greater will be the quantity of gas that can diffuse in a given period of time. In some pulmonary diseases, such as *emphysema,* large portions of the lungs may be destroyed and the total area of the pulmonary membrane greatly decreased. Sometimes the decrease is sufficient to cause continual respiratory embarrassment.

3. The thinner the membrane, the greater will be the rate of gaseous diffusion. The normal pulmonary membrane is sufficiently thin so that venous blood entering a pulmonary capillary can be brought almost to complete gaseous equilibrium with the alveolar air within approximately one fifth of a second. Occasionally, though, the thickness of the membrane increases many-fold, especially when the lungs become edematous because of pulmonary congestion, pneumonia, or some other lung disease. The person may die simply because gases cannot diffuse through the membrane rapidly enough.

4. The diffusion coefficient of a gas is proportional to the amount of gas that dissolves in the fluid and is inversely proportional to the square root of the molecular weight. Therefore, highly soluble gases and gases with small molecular weights diffuse more rapidly than less soluble gases and gases with large molecular weights. If the diffusion coefficient of oxygen is considered to be 1, the relative rates of diffusion of the three important respiratory gases are the following:

Oxygen = 1
Carbon dioxide = 20.3
Nitrogen = 0.53

Thus carbon dioxide diffuses through the pulmonary membrane far more easily than either oxygen or nitrogen. The main reason for these differences in rate of diffusion is that carbon dioxide is about 20 times as soluble as oxygen in the fluids of the pulmonary membrane, while nitrogen is only about half as soluble as oxygen.

TRANSPORT OF OXYGEN THROUGH THE TISSUES

Diffusion of Oxygen from the Alveoli into the Blood and from the Blood into the Tissues

Figure 163 illustrates the overall transport of oxygen from the alveolus to the tissue cell, showing diffusion of oxygen from the alveolus into the pulmonary blood, transport of the blood through the arteries to the tissue capillaries, and, finally, diffusion of the oxygen from the capillaries to the cell.

The data of Table 4 explain why oxygen diffuses from the alveoli into the pulmonary blood. The oxygen pressure (Po_2) of unaerated blood entering the pulmonary capillaries is only 40 mm. Hg in comparison with the alveolar Po_2 of 104 mm. Hg. The diffusion gradient, therefore, is 64 mm. Hg, causing extremely rapid flow of oxygen into the blood. During the very short time that the blood remains in the capillary, only about one second, it attains a Po_2 of approximately 100 mm. Hg.

The aerated blood then flows from the lungs to the tissue capillaries, where oxygen diffuses from the blood into the tissue cells, in which the Po_2 always remains very low, the reason for which may be explained as follows: When oxygen enters a cell it reacts very readily with sugars, fat, and proteins to form carbon dioxide and water, as is illustrated in Figure 164. As a result, the concentration of oxygen in the cell rarely rises above 30 mm. Hg. The high pressure of the oxygen in the capillaries, 100 mm. Hg, causes it to diffuse into the interstitial fluid and through the cellular membrane to the interior of the cells.

In summary, oxygen movement, both into the blood of the lungs and from the blood to the tissues, is caused by diffusion. The pressure gradient from the lungs to the pulmonary blood is sufficient to keep oxygen always diffusing into the blood, and the pressure gradient from the capillary blood to the tissue cells is always sufficient to keep oxygen diffusing from the blood to the cells.

Transport of Oxygen in the Blood by Hemoglobin

When oxygen diffuses from the lungs into the blood a small proportion of it becomes dissolved

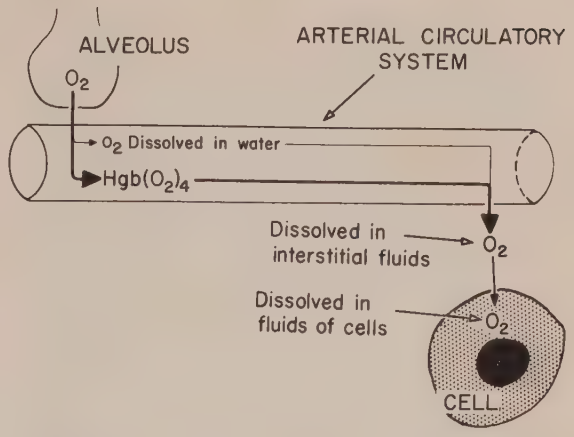


Figure 163. Transport of oxygen from the alveolus to the tissue cell.

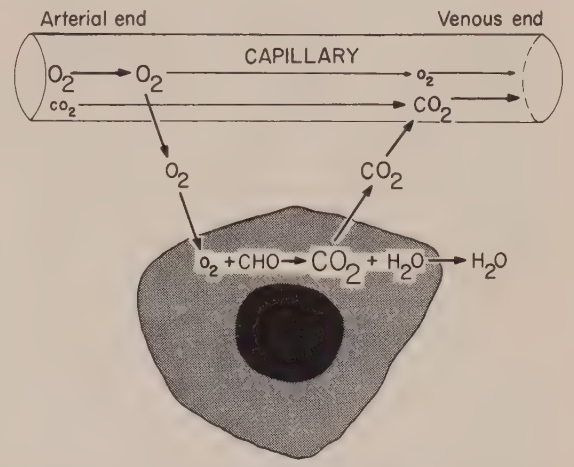


Figure 164. Diffusion of oxygen from the capillary to the tissue cell, formation of carbon dioxide inside the cell, and diffusion of carbon dioxide back to the tissue capillary.

in the fluids of the plasma and red cells, but approximately 60 times as much combines immediately with the hemoglobin of the red blood cells. Then, when the blood passes through the tissue capillaries, oxygen splits away from the hemoglobin and diffuses into the tissues. Thus, hemoglobin acts as a carrier of oxygen, increasing the amount of oxygen that can be transported

Table 4. *Relative Gaseous Pressures in the Alveoli and the Blood*

Gas	*Alveolar Air*	*Venous Blood*	*Arterial Blood*
O_2	104	40	100
CO_2	40	45	40
N_2	569	567	569

from the lungs to the tissues to some 60 times more than that which could possibly be transported only in the dissolved state.

Figure 165 illustrates the so-called *oxygen-hemoglobin dissociation* curve. This curve shows the per cent of the hemoglobin that is combined, or *saturated,* with oxygen at each oxygen pressure level. Aerated blood leaving the lungs usually has an oxygen pressure of about 100 mm. Hg. Referring to the curve, it is seen that at this pressure *approximately 97 per cent of the hemoglobin is combined with oxygen.* As the blood passes through the tissue capillaries, the oxygen pressure falls normally to about 40 mm. Hg. At this pressure *only about 70 per cent* of the hemoglobin is combined with oxygen. Thus, about 27 per cent of the hemoglobin normally loses its oxygen to the tissue cells. Then, flowing continually through the circulation, it again and again regains oxygen in the lungs and loses it to the tissue cells.

Utilization coefficient and the hemoglobin reserve capacity. The proportion of the hemoglobin that loses its oxygen to the tissues during each passage through the capillaries is called the *utilization coefficient.* In the normal person the utilization coefficient is 27 per cent, or, expressed another way, approximately one fourth of the hemoglobin is used to transport oxygen to the tissues under normal conditions. When the tissues are in extreme need of oxygen, the oxygen pressure in the tissues may fall to extremely low values, allowing oxygen to diffuse from the capillary blood much more rapidly than usual. As a result, the saturation of the hemoglobin can fall to as low as 20 per cent instead of the normal level of 70 per cent, and the utilization coefficient rises to 77 per cent. Therefore, without even increasing the rate of blood flow, the amount of oxygen transported to the tissues in times of serious need can be increased as much as three-fold. If one also remembers that the cardiac output can increase as much as six-fold in times of stress, then it is seen that the amount of oxygen that can be transported to the tissues can be increased to as much as 21 times normal, part of the increase being caused by an increase in the utilization coefficient and even more by the increased rate of blood flow.

Hemoglobin as an oxygen buffer. For cellular function to continue at a normal pace, the concentrations of all substances in the extracellular fluid must remain relatively constant at all times. One of the functions of hemoglobin is to keep the oxygen concentration in the tissues almost constant. This may be explained as follows: As blood flows through the capillaries, 27 per cent of the oxygen is usually removed from the hemoglobin, making the hemoglobin saturation fall to about 70 per cent. Referring again to Figure 165, it can be seen that to remove this much oxygen the oxygen pressure in the tissues can never rise above 45 mm. Hg. On the other hand, whenever the oxygen pressure falls to 20 mm. Hg, more than three fourths of the oxygen is released from the hemoglobin, and this rapid release usually keeps the tissue oxygen pressure from falling any lower. Thus, hemoglobin automatically releases oxygen to the tissues so that the tissue Po_2 remains almost always between the limits of 20 and 45 mm. Hg.

Carbon monoxide as a hemoglobin poison. Carbon monoxide combines with hemoglobin in almost identical manner as oxygen, except that the combination of carbon monoxide with hemoglobin is approximately 210 times as tenacious as the combination of oxygen with hemoglobin. Also, because the two substances combine with hemo-

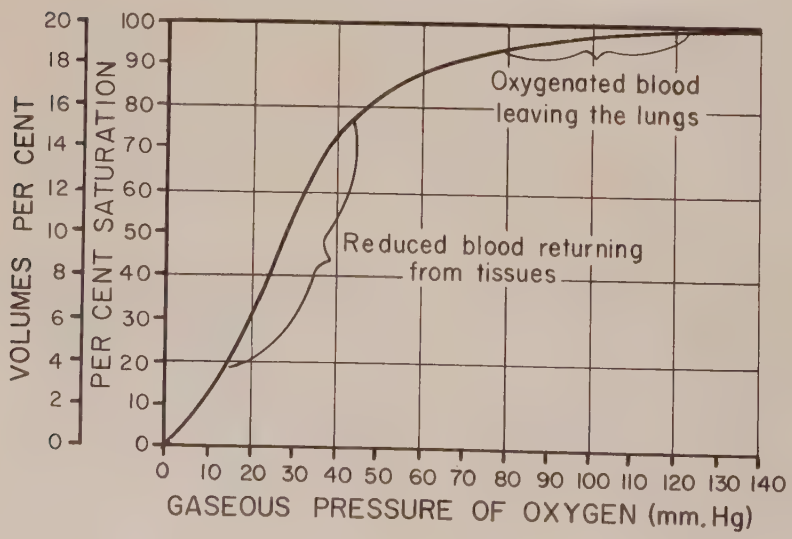


Figure 165. The oxygen-hemoglobin dissociation curve.

globin at the same point on the molecule, they both cannot be absorbed by the hemoglobin at the same time. Carbon monoxide mixed with air in a concentration of 0.1 per cent, 210 times less than that of oxygen, will cause half of the hemoglobin to combine with carbon monoxide and half with oxygen. When the concentration of carbon monoxide rises above approximately 0.2 per cent, which is still 100 times less than the normal concentration of oxygen, the quantity of hemoglobin available to transport oxygen becomes so slight that death ensues.

When death is about to occur from carbon monoxide poisoning, it can frequently be prevented by administering pure oxygen. The pure oxygen, on entering the alveoli, has a partial pressure of approximately 600 mm. Hg, six times the normal alveolar oxygen pressure. The force with which oxygen can combine with hemoglobin is increased six-fold, which forces the carbon monoxide off the hemoglobin six times as rapidly as would occur without treatment.

TRANSPORT OF CARBON DIOXIDE

Referring once again to Figure 164, it can be seen that when oxygen is used by the cells for metabolism carbon dioxide is formed. The pressure of carbon dioxide (P_{CO_2}) in the cells becomes very high, about 55 mm. Hg, and a pressure gradient develops between the cells and the blood in the capillary. This causes carbon dioxide to diffuse out of the cell, into the interstitial fluid, and from there into the capillary. As shown in Figure 166, it is transported in the blood to the lungs. The P_{CO_2} of the blood entering the tissue capillaries is about 40 mm. Hg, but this rises to 45 mm. Hg as carbon dioxide enters the blood from the cells. Therefore, the P_{CO_2} of the blood entering the lungs is approximately 45 mm. Hg, while the P_{CO_2} in the air of the alveolus is only 40 mm. Hg. A pressure gradient of about 5 mm. Hg exists between the blood and the alveolus, causing carbon dioxide to diffuse out of the blood. Furthermore, because of the extremely rapid diffusion coefficient of carbon dioxide, the P_{CO_2} of the pulmonary blood falls to almost complete equilibrium with the alveolar P_{CO_2}, that is, to 40 mm. Hg, before the blood leaves the capillary.

Chemical Combinations of Carbon Dioxide with the Blood

Only about 5 per cent of the carbon dioxide is transported in the blood in the dissolved state.

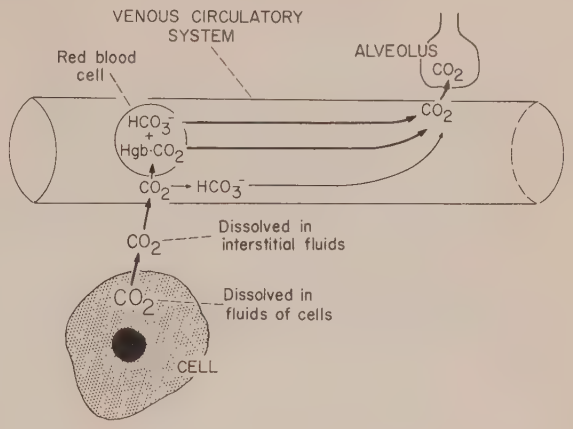


Figure 166. Transport of carbon dioxide from the tissue cell to the alveolus, illustrating some of the chemical combinations of carbon dioxide with blood.

About 95 per cent of the carbon dioxide diffuses from the plasma into the red cell where it undergoes two reactions: First, about 30 per cent of the total carbon dioxide entering the blood combines directly with hemoglobin to form a compound called *carbaminohemoglobin.* This reaction does not occur at the same point on the hemoglobin molecule as the reaction between oxygen and hemoglobin. Therefore, hemoglobin can combine with both carbon dioxide and oxygen at the same time. In this way hemoglobin acts as a carrier not only of oxygen but also of carbon dioxide.

The second chemical reaction of carbon dioxide in the red blood cells is with water to form carbonic acid. Red blood cells contain an enzyme called *carbonic anhydrase* which speeds this reaction approximately 250-fold. Therefore, very large quantities of carbon dioxide (about 65 per cent of that entering the blood) combine with water to form carbonic acid inside the red blood cells while only small quantities are formed in the plasma. The carbonic acid in the cells immediately reacts with the acid-base buffers of the cells and becomes mainly bicarbonate ion. This reaction prevents the acidity of the cell from becoming too great.

To recapitulate, carbon dioxide is transported in the blood in three principal forms. First, only a very small quantity is transported as dissolved carbon dioxide. Second, approximately 65 per cent is transported in the form of bicarbonate ion. And, third, approximately 30 per cent is transported in the form of carbaminohemoglobin.

When the blood enters the pulmonary capillaries all the chemical combinations of carbon dioxide with blood are reversed, and the carbon dioxide is released into the alveoli.

REFERENCES

Altman, P. L., Gibson, J. F., and Wang, C. C.: *Handbook of Respiration.* Philadelphia, W. B. Saunders Company, 1958.

Bouhuys, A., and Lundin, G.: Distribution of inspired gas in lungs. *Physiol. Rev.* 39:731, 1959.

Boyden, E. A.: *Segmental Anatomy of the Lungs: A Study of the Patterns of the Segmental Bronchi and Related Vessels.* New York, McGraw-Hill Book Company, Inc., 1955.

Campbell, E. J. M.: *The Respiratory Muscles.* Chicago, Year Book Publishers, Inc., 1958.

Comroe, J. H., Forster, R. E., Dubois, A. B., Briscoe, W. A., and Carlsen, E.: *The Lung.* 2nd Ed. Chicago, Year Book Publishers, Inc., 1962.

Du Brul, K. L.: *Evolution of the Speech Apparatus.* Springfield, Ill., Charles C Thomas, 1958.

Mead, J.: Mechanical properties of lungs. *Physiol. Rev.* 41:281, 1961.

Negus, V.: *Comparative Anatomy and Physiology of the Nose and Paranasal Sinuses.* Baltimore, The Williams & Wilkins Company, 1958.

Severinghaus, J. W.: Respiration. *Ann. Rev. Physiol.* 24:421, 1962.

Widdicombe, J. G.: Regulation of tracheobronchial smooth muscle. *Physiol. Rev.* 43:1, 1963.

Wylie, W. D.: *A Practice of Anesthesia.* Chicago, Year Book Publishers, Inc., 1960.

Wyss, O. A. M.: Respiration. *Ann. Rev. Physiol.* 25:143, 1963.

REGULATION OF RESPIRATION AND THE PHYSIOLOGY OF RESPIRATORY ABNORMALITIES

18

THE BASIC RHYTHM OF RESPIRATION

The nervous system regulates the rhythmic muscular contractions of breathing. Respiratory impulses begin in the *respiratory center* located in the recticular substance of the medulla and pons, as illustrated in Figure 167, and are transmitted down the spinal cord to the appropriate muscles of respiration. The *diaphragm*, the most important respiratory muscle, receives its stimuli through a special nerve, the *phrenic nerve*, which leaves the spinal cord in the neck and passes downward through the chest to reach the diaphragm.

Rhythmic Oscillation in the Respiratory Center

The rhythmic respiratory cycle is caused by continuous oscillation in the respiratory center. Anteriorly in the medulla is the so-called *inspiratory center*, which causes the inspiratory muscles to contract. Immediately behind and slightly

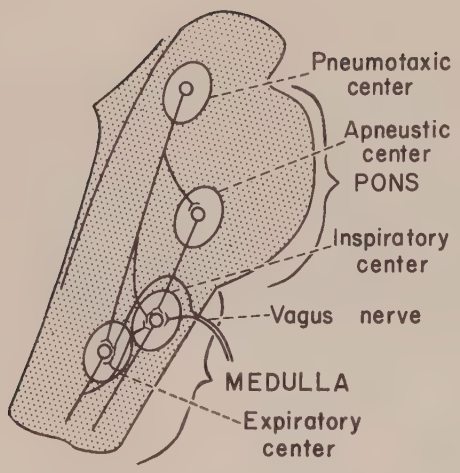


Figure 167. The respiratory center.

above is the *expiratory center,* which causes the expiratory muscles to contract. In the lower part of the pons is the *apneustic center,* which, when stimulated, causes very prolonged inspiration punctuated by occasional expiratory gasps. In the top of the pons is the *pneumotaxic center,* which inhibits the apneustic type of breathing and also accelerates the rate of respiration.

The most generally accepted theory of operation of the respiratory center is the following: It is believed that the inspiratory center is composed of neurons that, when not inhibited, transmit a continuous stream of impulses to the inspiratory muscles and thereby cause inspiration. However, at the same time that these inspiratory impulses are transmitted, others are also transmitted through a *delay circuit* to the expiratory center, exciting this two to three seconds later. The expiratory center then excites the expiratory muscles and simultaneously inhibits the inspiratory center, allowing the inspiratory muscles to relax. After another two to three seconds, the impulses of the expiratory center die out, allowing the inspiratory center to become active again and to initiate a new cycle of respiration.

The basic oscillation occurring between the inspiratory and expiratory centers of the medulla causes only weak signals to be transmitted to the respiratory muscles, but the apneustic center makes the signals to the inspiratory muscles become powerful, causing forceful inspiration. The pneumotaxic center makes the expiratory signals strong, and makes the rhythm of respiration regular and smooth. Thus, the apneustic and pneumotaxic centers enhance the activities of the medullary centers to provide deeper respiration, and they help to control the ventilation in proportion to the body's need for respiration.

Effect of the Hering-Breuer reflex on the basic rhythm of respiration. Located throughout the lungs are nerve endings that become stimulated when the lungs are distended. Impulses from these pass through the vagus nerve to the medulla, where they inhibit the inspiratory center and excite the expiratory center. One of the major functions of this reflex, called the *Hering-Breuer reflex,* is to prevent overinflation of the lungs. Excessive distention of the lungs can damage them by tearing the tissues, or, more frequently, it temporarily blocks the flow of blood into the heart because of dilatation of the lung vessels which take up the blood instead.

The Hering-Breuer reflex also aids in maintaining the basic rhythm of respiration. As the lungs distend during inspiration this reflex inhibits the inspiratory center, stopping further inspiration, and it simultaneously excites the expiratory center to promote expiration. Then, as the lungs deflate, the Hering-Breuer reflex ceases, allowing the inspiratory center to become excited again and the expiratory center to become inhibited. Therefore, this reflex helps to maintain the alternating rhythmic stimulation of the inspiratory and expiratory centers.

It is evident, then, that at least two different feedback mechanisms help to maintain the respiratory cycle, one intrinsic in the medulla and one operating extrinsically through the Hering-Breuer reflex. If either one of them becomes nonfunctional, the other is still capable of continuing the cycle. This duplication of neurogenic mechanisms is characteristic of most of the important functions of the brain, for any system that must continue to operate without a single instance of failure for up to 100 years must have a large margin of safety should some portion of the system become temporarily damaged.

Failure of the respiratory center. Occasionally all the oscillating mechanisms of the respiratory center fail at the same time. One of the more frequent causes of this is cerebral concussion or some other intracerebral abnormality that causes excess pressure on the medulla. The pressure collapses the blood vessels supplying the respiratory center, which blocks all activity of the medulla and as a result stops the respiratory rhythm. Another common cause of failure is acute poliomyelitis, which sometimes destroys neuronal cells in the recticular substance of the hindbrain, thereby depressing the respiratory center. Finally, the most frequent of all causes of respiratory failure is attempted suicide with sleeping drugs. These anesthetize the neuronal cells and in this way stop the rhythm of respiration.

Failure of the respiratory rhythm is one of the most difficult of all abnormalities of the body to

overcome. In general, very few drugs can be used to excite the respiratory center, and those that are available—caffeine, picrotoxin, Coramine, and a few others—are so weak in their effects on the respiratory center that they are of little value. In general, the only effective treatment is to give the patient prolonged artificial respiration in the hope that the cause of the respiratory depression will be rectified by the body itself. For instance, respiratory depression caused by poliomyelitis, pressure on the brain, or sleeping drugs is usually reversible provided artificial respiration is maintained long enough.

REGULATION OF ALVEOLAR VENTILATION

Even though the basic rhythm of respiration is continuous, the rate of rhythm and the depth of respiration vary tremendously in response to varying physiologic conditions in the body. Obviously, the more oxygen needed by the body and the more carbon dioxide the body needs to expel, the greater must be the alveolar ventilation.

Effect of Carbon Dioxide on Alveolar Ventilation

The most powerful stimulus known to affect the respiratory center is *carbon dioxide.* When the carbon dioxide content of the blood increases above normal, all portions of the respiratory center become excited. The intensity of both inspiratory and expiratory signals becomes greatly enhanced, and the rapidity of the oscillation back and forth between the inspiratory and expiratory centers is also speeded up. Therefore, *both the rate and depth of respiration are increased.* This effect is illustrated in Figure 168, which shows that increased carbon dioxide in the blood can increase the ventilatory rate to as much as 10 times normal.

Importance of carbon dioxide as a stimulus to respiration. The concentration of carbon dioxide in the body fluids is mainly dependent on the alveolar ventilatory rate. When alveolar ventilation decreases, the alveolar concentration of carbon dioxide increases, which causes the concentration of carbon dioxide in all the body fluids to increase too. On the other hand, when alveolar ventilation increases, the alveolar concentration of carbon dioxide decreases and so does the concentration of carbon dioxide in all the body fluids. In fact, there is no other means except changing alveolar ventilation for controlling the carbon dioxide concentration in the body fluids. Also, because carbon dioxide is one of the major end products of metabolism, its concentration in the tissues has a profound influence on the rates of cellular metabolic reactions. An excess of carbon dioxide slows up the reactions, while a diminished quantity enhances the chemical reactions. Therefore, it is quite important that alveolar ventilation be controlled by the carbon dioxide in the body so that the carbon dioxide concentration can in turn be maintained constant.

Regulation of Alveolar Ventilation by the Hydrogen Ion Concentration

The second most powerful influence on alveolar ventilation is the hydrogen ion concentration in the body fluids. When the hydrogen ion concentration becomes high—that is, when the pH

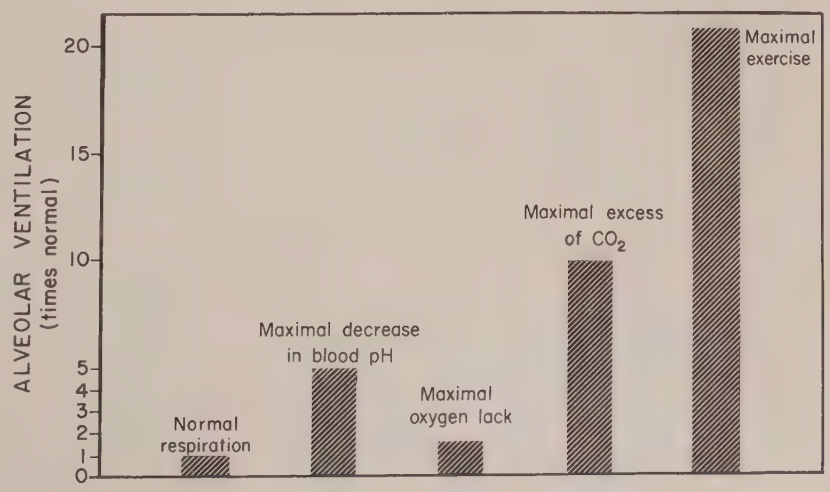


Figure 168. Effect on alveolar ventilation of maximal decrease in blood pH, maximal oxygen lack, maximal excess of CO_2, and maximal exercise.

becomes low—alveolar ventilation is increased. However, the maximum amount that ventilation can be increased by increasing the hydrogen ion concentration is about five-fold. Therefore, an increase in the hydrogen ion concentration is only one half as potent a stimulus of respiration as an increase in carbon dioxide concentration.

Importance of hydrogen ion concentration as a means for regulating alveolar ventilation. Regulation of respiration by the hydrogen ion concentration helps to maintain normal acid-base balance in the body fluids, which was discussed in more detail in Chapter 8. When carbon dioxide is blown off through the lungs, loss of this carbon dioxide from the body fluids shifts the chemical equilibria of the acid-base buffers in such a way that much of the carbonic acid of the body fluids dissociates into water and carbon dioxide. This removes some of the body's acid and thereby decreases the hydrogen ion concentration. Therefore, stimulation of the respiratory center by increased hydrogen ion concentration sets off a feedback mechanism that automatically decreases the hydrogen ion concentration back toward normal. Conversely, decreased hydrogen ion concentration decreases alveolar ventilation, thus increasing the acidity of the fluids back toward normal.

Regulation of Alveolar Ventilation by Oxygen Deficiency

On first thought most students have the idea that respiration should be regulated mainly by the needs of the body for oxygen. However, at ordinary ventilatory rates the hemoglobin always becomes almost completely saturated with oxygen, and extreme increases or moderate decreases in alveolar ventilation make very little difference in the amount of oxygen carried by the hemoglobin. Therefore, there is no need for very acute regulation of respiration to maintain constant oxygen concentration in the blood.

On rare occasions, however, the oxygen concentration in the alveoli does fall too low to supply adequate quantities of oxygen to the hemoglobin. This occurs especially when one ascends to high altitudes where the oxygen concentration in the atmosphere is very low. Or it occurs when a person contracts pneumonia or some disease that reduces the oxygen in the alveoli. Under these conditions the respiratory system needs to be stimulated by *oxygen deficiency*. A mechanism for this is the *chemoreceptor system*, shown in Figure 169. Small *aortic* and *carotid bodies* lie adjacent to the aorta and carotid arteries in the chest and neck, each of these having an abundant arterial blood supply, and containing ganglion-like cells called *chemoreceptors*, which are sensitive to the lack of oxygen in the blood. When stimulated, these receptors send impulses along the vagus and glossopharyngeal nerves to the medulla, where they stimulate the respiratory center to increase alveolar ventilation.

The chemoreceptor system is not a powerful stimulator of respiration in comparison with the stimulation of the respiratory center by excess carbon dioxide or hydrogen ions. As illustrated in Figure 168, maximum oxygen lack can increase alveolar ventilation to approximately 1⅔ times its normal value in comparison with an increase of 10 times caused by excess carbon dioxide and 5 times caused by excess hydrogen ions.

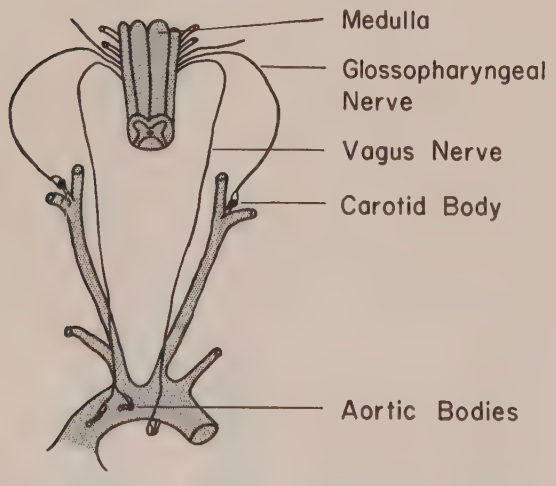


Figure 169. The chemoreceptor system for stimulating respiration.

Effect of Exercise on Alveolar Ventilation

Alveolar ventilation increases almost directly in proportion to the amount of work performed by the body during exercise, reaching as high as 100 to 120 liters per minute in the most strenuous exercise. This value is about 20 times that during normal quiet respiration, as illustrated in Figure 168.

Two separate theories have been proposed in an attempt to explain the extreme effectiveness of respiratory control in exercise. First, it has been claimed that the motor portions of the cerebral cortex send impulses directly to the respiratory center at the same time that they send impulses to the muscles, exciting the respiratory center and thereby increasing respiration along with the increase in muscle activity. Second, it has been

claimed that chemoreceptors similar to those in the aortic and carotid bodies are also present in the walls of the great veins or pulmonary artery. These chemoreceptors are supposedly exposed to the extremely high carbon dioxide concentration and the extremely low oxygen concentration in the venous blood. Obviously, they would be strongly stimulated during exercise and could cause greatly increased respiration. Unfortunately, there is at present insufficient research evidence to prove or disprove either of these two theories; it is very possible, if not likely, that both of the mechanisms operate in exercise.

Other Factors That Affect Alveolar Ventilation

Effect of arterial pressure. Arterial pressure is another factor that helps regulate alveolar ventilation. This operates through the *pressoreceptor system* which was described in Chapter 15. When the arterial pressure is high, the respiratory center is depressed, and ventilation is reduced correspondingly. Conversely, when the blood pressure is low and blood flow to the tissues is poor, ventilation is increased; this partially compensates for the poor blood flow by allowing better oxygenation and better removal of carbon dioxide in the lungs.

Effect of psychic stimulation. Impulses initiated by psychic stimulation of the cerebral cortex can also affect respiration. For instance, anxiety states frequently can lead to intense hyperventilation, sometimes so much so that the person makes his body fluids alkalotic, thus precipitating tetany.

Effect of sensory impulses. Sensory impulses from all parts of the body can affect respiration. The effect of entering a cold shower is well known; this causes an intense inspiratory gasp followed by a period of prolonged inspiration and then by rapid, forceful breathing. Even a pin prick can cause sudden changes in the rate and depth of respiration. However, these factors affect alveolar ventilation only temporarily, for the chemical factors mentioned previously soon become dominant over these aberrant effects and suppress them.

Effect of speech. The speech centers of the brain also control respiration at times. When one talks it is as important to control the flow of air between the vocal cords as to control the vocal cords themselves. Therefore, whenever nerve impulses are transmitted from the brain to the vocal cords, collateral impulses are sent simultaneously into the respiratory system.

The Multiple Factor Theory for Control of Respiration

It is obvious from the preceding discussions that no single factor regulates the respiratory centers, but that many different factors, including carbon dioxide concentration, hydrogen ion concentration, oxygen availability, exercise, arterial pressure, sensory phenomena, and speech all help to regulate respiration. The most generally accepted theory for the regulation of respiration, then, is the so-called *multiple factor theory.* Some factors are known to be more important than others—carbon dioxide or exercise, for example. Others are of immediate importance, such as hydrogen ion concentration, and others are of relatively minor importance, such as oxygen deficiency and arterial pressure.

PHYSIOLOGY OF RESPIRATORY ABNORMALITIES

Anoxia

One of the most important effects of most respiratory diseases is anoxia, which means diminished availability of oxygen to the cells of the body. The different types of anoxia that can occur are anoxic anoxia, stagnant anoxia, anemic anoxia, and histotoxic anoxia, the principal mechanisms of which are shown in Figure 170.

Anoxic anoxia. Anoxic anoxia means failure of oxygen to reach the blood of the lungs. The obvious causes of anoxic anoxia are (1) *too little oxygen in the atmosphere,* (2) *obstruction of the respiratory passages,* (3) *thickening of the pulmonary membrane,* and (4) *decreased area of the pulmonary membrane.*

Stagnant anoxia. Stagnant anoxia means failure to transport adequate oxygen to the tissues because of slow blood flow. The most common cause of stagnant anoxia is low cardiac output caused by *heart failure.* Immediately after a heart attack the blood flow to the tissues may be so slight that the person may die of the stagnant anoxia itself. At other times the transport of oxygen will be barely enough to keep him alive but his tissues nevertheless suffer severely from oxygen deficiency.

Anemic anoxia. Anemic anoxia means too little hemoglobin in the blood to transport oxygen to the tissues. This may be caused by any one of three different abnormalities: First, the person may have *anemia,* which means too few red blood cells. Second, he may have a sufficient number of red blood cells but *too little hemoglobin in the*

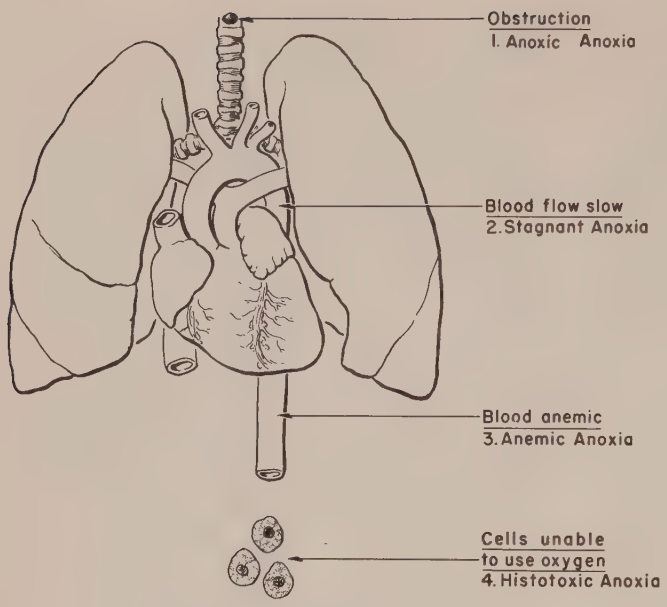


Figure 170. The types of anoxia.

cells. Third, he may have plenty of hemoglobin, but much of it may have been *poisoned by carbon monoxide* or some other poison so that it cannot transport oxygen; for instance, breathing air with a content of only 0.2 per cent carbon monoxide can decrease the available hemoglobin to as low as one-third normal.

Histotoxic anoxia. This means failure of the tissues to utilize oxygen even though adequate quantities are transported to them. The classic cause of serious histotoxic anoxia is *cyanide poisoning;* this blocks the enzymes responsible for oxidation in the cell. Mild forms of histotoxic anoxia frequently occur with *vitamin deficiencies,* for lack of certain of the vitamins often results in diminished quantities of oxidative enzymes in the cells.

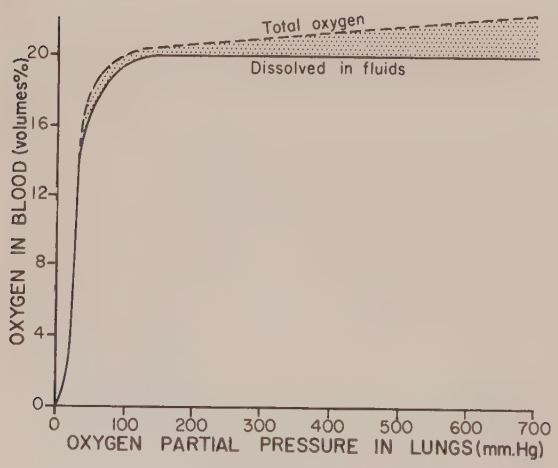


Figure 171. Effect of elevated alveolar partial pressure of oxygen on the total oxygen transported by each 100 ml. of blood.

Oxygen Therapy

In most types of anoxia, therapy with purified oxygen can be tremendously beneficial, the benefit occurring in three separate ways, (a) by increasing the alveolar oxygen pressure, which increases the rate of diffusion through the pulmonary membrane, (b) by increasing the quantity of oxygen dissolved in the fluids of the blood in addition to that combined with the hemoglobin, and (c) by decreasing the volume of gases that must flow through the trachea to aerate the alveoli.

The normal partial pressure of oxygen in the alveoli is about 104 mm. Hg, but when a person breathes pure oxygen, this can rise to as high as 600 mm. Hg, which can increase as much as sixfold the pressure gradient for oxygen diffusion through the pulmonary membrane.

Even when there is insufficient hemoglobin in the blood to carry enough oxygen to the tissues, oxygen therapy can still be beneficial. The reason for this is the following: Ordinarily, almost no oxygen is dissolved in the fluids of the blood, but, instead, it combines with the hemoglobin and is transported to the tissues in this form. Yet when the partial pressure of oxygen in the alveoli becomes 600 mm. Hg, 2 ml. of oxygen becomes dissolved in the fluid of every 100 ml. of blood and is transported in this form to the tissues. This effect is illustrated in Figure 171. The lower solid

curve of this figure shows the total quantity of oxygen combined with hemoglobin at different oxygen partial pressures, and the upper curve is the total oxygen in the blood including that in the dissolved state. At a normal alveolar oxygen partial pressure of about 100 mm. Hg, almost no oxygen is dissolved in the fluids. Yet as the alveolar oxygen partial pressure rises to 600 mm. Hg, one sees by the shaded area of the figure that reasonable quantities of oxygen begin to dissolve in the fluids. This oxygen is the difference between life and death in some anoxic persons.

The third means by which oxygen therapy can benefit some types of anoxia, is by decreasing the volume of gases that must flow in and out of the lungs through the trachea. If the respiratory passageways are partially obstructed, the blood can still be adequately oxygenated with only one-fifth normal alveolar ventilation if the person breathes pure oxygen rather than air. The reason for this is that the concentration of oxygen in air is only 20 per cent. Yet the decreased alveolar ventilation will still cause a buildup of carbon dioxide in the body fluids, sometimes enough to produce serious acidosis even though the tissues are adequately oxygenated.

Physiology of oxygen therapy in the different types of anoxia. Keeping the preceding principles of oxygen therapy in mind, the value of oxygen in the different types of anoxia may be summarized as follows:

In *anoxic anoxia,* the amount of oxygen made available to the blood can usually be increased 500 per cent or more with oxygen therapy, which means that a person with pulmonary disease that allows only one-sixth normal oxygenation can still be kept alive with oxygen therapy.

In *anemic anoxia,* oxygen therapy is moderately valuable. The rate of blood flow through the tissues is often as much as three times normal, and the extra oxygen dissolved in the fluids of this fast flowing blood can sometimes supply as much as 40 per cent of the oxygen needs of the body, thus making up for diminished availability of hemoglobin to transport oxygen. Therefore, persons with such severe anemia that only 60 per cent of the oxygen needs are being supplied by normal transport of oxygen can still be kept alive with oxygen therapy.

In *stagnant anoxia,* oxygen therapy is only mildly beneficial because the anoxia is caused by stagnation of blood flow rather than by lack of adequate oxygen in the blood. Yet even in this condition oxygen therapy can increase the amount of oxygen transported to the tissues about 13 per cent, which may mean the difference between life and death. For instance, patients with severe heart attacks almost always receive at least some degree of comfort from oxygen therapy.

In *histotoxic anoxia,* oxygen therapy is generally of almost no benefit, for the tissues are already supplied with more than adequate quantities of oxygen, but they simply cannot utilize it.

Dyspnea

Dyspnea means *air hunger* or, in other words, a feeling that more ventilation of the blood is needed than is being attained. Most instances of dyspnea occur when some respiratory abnormality causes the blood to become anoxic and to collect too much carbon dioxide. The respiratory center becomes overly excited, and impulses are transmitted to the conscious portions of the brain apprising the psyche of the need for increased ventilation. This causes a feeling of air hunger or dyspnea.

However, an occasional person develops *psychic dyspnea* because of a neurosis. His blood has normal carbon dioxide and is adequately oxygenated, but the neurosis causes him to become unduly conscious of his respiration, making him feel that he is not receiving adequate quantities of air. One of the neuroses that most frequently causes this type of dyspnea is cardiac neurosis, for many who fear cardiac disease have known cardiac patients who have had continuous dyspnea because of stagnant anoxia. Therefore, the neurotic who believes that he has heart disease often develops psychic dyspnea.

Pneumonia

Pneumonia is an infectious lung disease that causes the walls of the alveoli to become inflamed and edematous and the spaces in the alveoli to become filled with fluid and blood cells. This effect is illustrated in the center of Figure 172. Pneumonia causes anoxic anoxia for two reasons: first, because fluid and blood cells fill many alveoli, and, second, because the membranes of those alveoli that are still aerated are frequently so thickened with edema that oxygen cannot diffuse with ease.

In pneumonia, as well as in most other respiratory diseases, carbon dioxide is usually excreted rapidly enough not to cause serious acidosis, because carbon dioxide diffuses through the pulmonary membrane approximately 20 times as easily as oxygen. For this reason pulmonary diseases usually depress oxygenation of the blood many times more than they do the excretion of carbon dioxide.

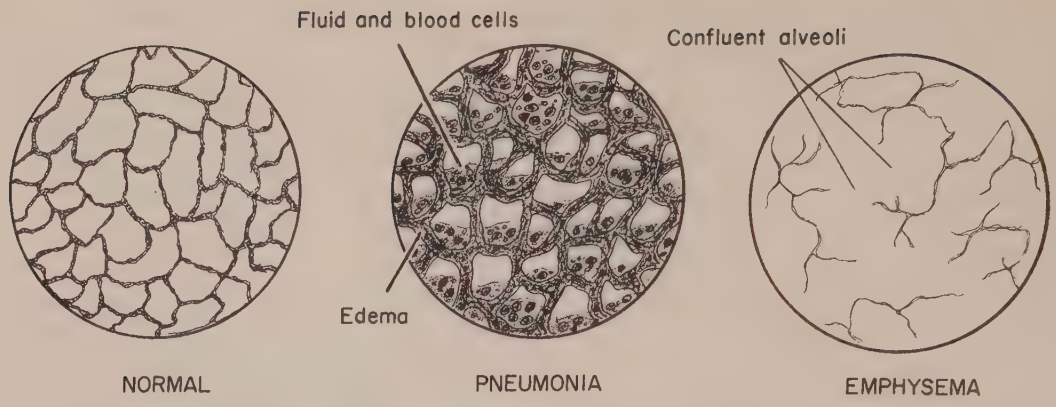


Figure 172. Histologic appearance of the normal lung, of the lung with pneumonia, and of the emphysematous lung.

Pulmonary Edema

Pulmonary edema means collection of fluid in the interstitial spaces of the lungs and in the alveoli; this affects respiration in the same way pneumonia affects it. Generalized pulmonary edema is usually caused by failure of the left heart to pump blood from the pulmonary circulation into the systemic circulation. This can result from mitral or aortic valvular disease or from failure of the left ventricular muscle. In any event, blood is dammed up in the pulmonary circulation, and the pulmonary capillary pressure rises. When this pressure becomes greater than the colloid osmotic pressure of the blood, 28 mm. Hg, fluid transudes very rapidly from the plasma into the alveoli and interstitial spaces of the lungs, causing anoxic anoxia. Sometimes acute failure of the left heart causes congestion to occur so rapidly that the person dies of anoxia in only 20 to 40 minutes.

Emphysema

In emphysema, a disease caused most frequently by chronic bronchial infection resulting from smoking, large portions of the alveolar walls have degenerated or have been destroyed; this is illustrated in Figure 172. The total surface area of the pulmonary membrane becomes greatly diminished, which also diminishes the aeration of the blood. Anoxic anoxia occurs, and the quantity of carbon dioxide in the body fluids becomes mildly increased.

The emphysematous person also usually develops pulmonary hypertension, for every time the wall of an alveolus is destroyed the blood vessels in the wall are also destroyed. This increases the pulmonary resistance, which in turn elevates the pulmonary arterial pressure.

Atelectasis

Atelectasis means collapse of a portion of the lung or of an entire lung. A common cause of atelectasis is a chest wound that allows air to leak into the pleural cavity. The surface tension of the fluid in the alveoli and the elastic fibers in the interstitial spaces of the lungs cause the lungs to collapse to a small size, the alveoli losing all their air. Another common cause of atelectasis is plugging of a bronchus, in which case the air in the alveoli is absorbed into the blood, thus collapsing the alveoli.

Not only do the alveoli collapse in atelectasis, but the blood vessels also constrict because of the alveolar anoxia that supervenes, which has a direct constricting effect on the pulmonary arteries. Consequently, blood flow through the collapsed lung decreases greatly, allowing the major portion of the pulmonary blood to flow through areas of the lungs that are still aerated. Because of this shift of blood flow, atelectasis of half of a lung or even of a whole lung often does not greatly diminish the aeration of the blood.

Asthma

Asthma is usually caused by an allergic reaction to pollens in the air. The reaction causes spasm of the bronchioles, thus impeding air flow into and out of the lungs. For reasons that are not too clearly understood, the outward flow of air through the bronchioles is obstructed more than the inflow, which allows the asthmatic person to inspire with ease but to expire with difficulty. As a result, the lungs become progressively more and more distended, and with repeated asthmatic attacks year in and year out, the prolonged distention of the thoracic cage causes the chest to

become barrel shaped. Asthma rarely becomes severe enough to cause serious anoxia, but it does cause severe dyspnea. Asthma can usually be treated by administering drugs, especially epinephrine, that relax the bronchiolar musculature. Also, it can often be prevented by desensitization to the causative pollens, as discussed in Chapter 10.

Absorption of Air Entrapped in Different Areas of the Body

Though the entrapping of air in different areas of the body is not a respiratory disorder, the principles applicable to respiration are also applicable to absorption of this air. Figure 173 illustrates to the left a large bubble of air containing normal quantities of oxygen, nitrogen, and water vapor. This air bubble has become trapped somewhere in the body's tissues. The oxygen pressure in the surrounding interstitial fluid is quite low because of continual utilization of oxygen by the cells. Therefore, oxygen begins to be absorbed from the bubble very rapidly. The bubble becomes smaller; the ratio of nitrogen to oxygen in the bubble becomes progressively greater. As this happens, the pressure of nitrogen in the bubble becomes greater than the nitrogen pressure in the interstitial fluid, and nitrogen, too, begins to be absorbed. The water vapor always maintains a pressure of exactly 47 mm. Hg as long as the body temperature remains constant, and carbon dioxide diffuses into the bubble from the tissues. When the bubble becomes smaller the water vapor simply condenses on the walls of the bubble, and the carbon dioxide diffuses back into the tissues. Eventually, all of the oxygen and nitrogen will have been absorbed, and the water vapor and carbon dioxide also will have disappeared into the interstitial fluids.

These principles of gas absorption from tissue spaces apply to air injected underneath the skin, air swallowed into the gastrointestinal tract, air that inadvertently leaks into the intrapleural space, or air anywhere in the body in contact with the interstitial fluids. These principles explain why the internal areas of the body, except the respiratory passageways and areas where gases are being formed too rapidly for all to be absorbed—the gut, for instance—are all devoid of gas spaces.

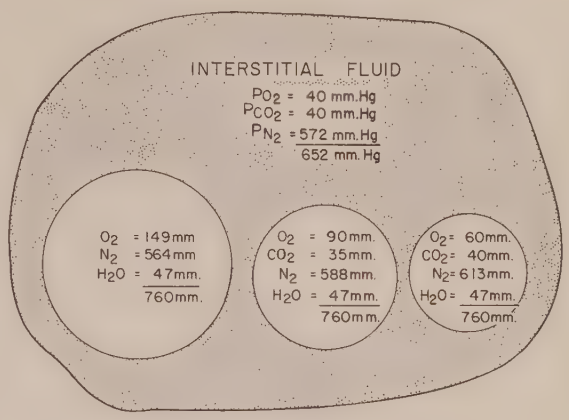


Figure 173. Absorption of air from the tissues.

REFERENCES

Cohen, M. I., and Wang, S. C.: Respiratory neuronal activity in pons of cat. *J. Neurophysiol.* 22:33, 1959.

Comroe, J. H., Jr., and Dripps, R. D.: *The Physiological Basis for Oxygen Therapy.* Springfield, Ill., Charles C Thomas, 1951.

Dawes, G. S., and Comroe, J. H., Jr.: Chemoreflexes from the heart and lungs. *Physiol. Rev.* 34:167, 1954.

Dejours, P.: Chemoreflexes in breathing. *Physiol. Rev.* 42:335, 1962.

Fishman, A. P.: Respiratory gases in the regulation of the pulmonary circulation. *Physiol. Rev.* 41:214, 1961.

Gray, J. S.: *Pulmonary Ventilation and Its Physiological Regulation.* Springfield, Ill., Charles C Thomas, 1950.

Grodins, F. S.: Analysis of factors concerned in regulation of breathing in exercise. *Physiol. Rev.* 30:220, 1950.

Rosebury, T.: *Experimental Air-Borne Infection.* Baltimore, The Williams & Wilkins Company, 1947.

Salmoiraghi, G. C., and Burns, B. D.: Notes on mechanism of rhythmic respiration. *J. Neurophysiol.* 23:14, 1960.

AVIATION, SPACE, AND DEEP SEA DIVING PHYSIOLOGY

19

One of the most important problems in aviation, space, and deep sea diving physiology is the altered barometric pressure to which the aviator, spaceman, or diver is exposed. Therefore, many of the physical principles applicable to respiration are also applicable to aviation, space flights, or dives deep beneath the sea.

Problems also arise in aviation, space, and diving physiology because of mechanical stresses on the body or because of extremes of climate. For instance, extreme acceleration during takeoff of a spaceship can cause the spaceman to weigh as much as 1500 pounds, a force that the body can withstand only when in the proper position. Also, in the upper atmosphere the air temperature is 20° to 50°C. below zero, and deep beneath the sea, pressures on the outside of the body can sometimes become so intense that the entire chest wall actually collapses.

EFFECTS OF LOW BAROMETRIC PRESSURE

Oxygen Deficiency at High Altitudes

At high altitudes the barometric pressure is low, as shown in Table 5, and the partial pressure of oxygen is reduced correspondingly. This decreases

the amount of oxygen absorbed into the blood, and leads to anoxia. Table 5 also shows the effect of low oxygen partial pressure on the saturation of arterial hemoglobin in the blood. For instance, at an altitude of approximately 23,000 feet only half of the arterial hemoglobin is saturated with oxygen. Obviously, this decreases the efficiency of oxygen transport to the tissues at least 50 per cent, and a person at this altitude will be debilitated because of tissue anoxia.

It will be noted also from Table 5 that the alveolar oxygen partial pressure at an altitude of 50,000 feet is only 1 mm. Hg and that the arterial oxygen saturation is only 1 per cent. Therefore, at this altitude, the small quantities of oxygen stored in the tissues actually diffuse backward into the blood and then from the blood into the lungs. Obviously, a person would have no more than a few minutes to live under these conditions.

Effects of oxygen deficiency. Figure 174 shows graphically the time required at various altitudes for oxygen deficiency to cause either collapse or coma. Above 30,000 feet a normal person lapses into coma in about one minute. At 20,000 feet the person usually does not go into coma, but after 10 or more minutes he exhibits signs of collapse such as weakness, mental haziness, and other deficiencies. He usually does not pass into coma until he ascends to altitudes of 22,000 to 24,000 feet, which is called his *ceiling*.

The earliest symptom of oxygen deficiency occurs in vision. Even at altitudes as low as 6000 to 10,000 feet, the sensitivity of the eyes to light begins to decrease—not enough to be noticed usually, but nevertheless enough to be measured by instruments. When the aviator rises to altitudes of 12,000 to 15,000 feet he is likely to experience alterations in psychic behavior. Some aviators tend to go to sleep while others develop euphoric exaltation. In almost all instances the acuity of reasoning becomes greatly depressed. In short, the aviator at this altitude is likely to develop a sleepy slap-happiness, and as he rises to 18,000 to 24,000 feet these symptoms often become so severe that control of the aircraft is endangered. For this reason safety regulations require that an aviator breathe oxygen when he ascends to dangerous altitudes. On occasion, he rises too high before applying his oxygen, and by that time his mental acuity may have become so depressed that he no longer knows his need for oxygen. This often results in a vicious cycle of depressed mentality, for

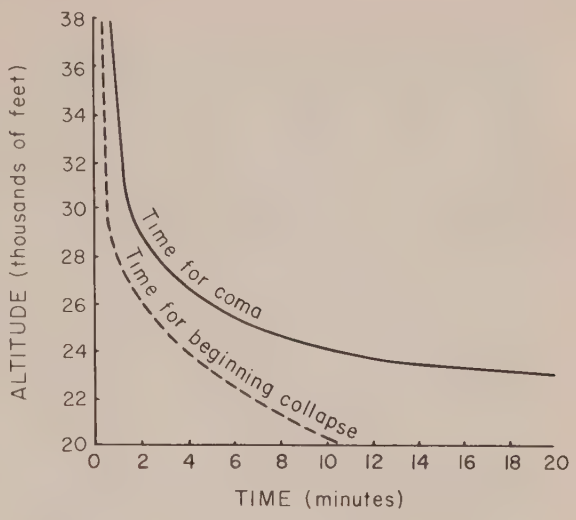


Figure 174. Time required at different altitudes for oxygen deficiency to cause collapse or coma. (From Armstrong: *Principles and Practice of Aviation Medicine.* The Williams & Wilkins Company.)

Table 5. *Effects of Low Atmospheric Pressures on Alveolar Oxygen Concentrations and on Arterial Oxygen Saturation*

Altitude (ft.)	*Barometric Pressure (mm. Hg)*	P_{O_2} *in Air (mm. Hg)*	BREATHING AIR		BREATHING PURE OXYGEN	
			P_{O_2} *in Alveoli (mm. Hg)*	*Arterial Oxygen Saturation (%)*	P_{O_2} *in Alveoli (mm. Hg)*	*Arterial Oxygen Saturation (%)*
0	760	159	104	97	673	100
10,000	523	110	67	90	436	100
20,000	349	73	40	70	262	100
30,000	226	47	21	20	139	99
40,000	141	29	8	5	58	87
50,000	87	18	1	1	16	15

he rises still higher without knowing what he is doing, his mentality becomes further depressed, he rises higher, and so forth until coma develops.

Effect of water vapor and carbon dioxide on alveolar oxygen partial pressure. Were it not for water vapor and carbon dioxide in the alveoli, the aviator could go to considerably higher heights than usual before developing oxygen deficiency. Regardless of the altitude, the alveolar pressure of water vapor does not change at all, and the pressure of carbon dioxide does not change greatly. Therefore, unless the barometric pressure is considerably greater than the combined pressures of these two gases, little space will be left in the alveoli for other gases. Figure 175 shows this effect. At sea level the combined pressures of all the gases in the alveoli is 760 mm. Hg. Carbon dioxide is continually excreted from the blood into the alveolar air, keeping the P_{CO_2} at a level of 40 mm. Hg, and water vapor continually evaporates from the surfaces of the alveoli creating a water vapor pressure of 47 mm. Hg, which remains constant as long as the body temperature is normal regardless of changes in barometric pressure. Thus, the combined pressure of water vapor and carbon dioxide at sea level is 40 plus 47 or 87 mm. Hg. Subtracting this from 760 mm. Hg, the remaining partial pressure available in the alveolus for both nitrogen and oxygen is 673 mm. Hg. This is more than adequate, so that the oxygen partial pressure is high enough to keep the arterial blood saturated with oxygen.

Now, let us see what effects the water vapor pressure and carbon dioxide partial pressure have on alveolar function at an altitude of 50,000 feet. Here, the total barometric pressure is only 87 mm. Hg. The water vapor pressure remains 47 mm. Hg because the body temperature does not change. The P_{CO_2} has fallen to about 24 mm. Hg because of increased respiration. Therefore, the combined pressure of water vapor and carbon dioxide is now 71 mm. Hg, which when subtracted from 87 mm. Hg leaves a total partial pressure for both nitrogen and oxygen of only 16 mm. Hg. Furthermore, the person is so anoxic that his blood absorbs oxygen out of the alveoli almost as rapidly as it enters. Therefore, the P_{O_2} falls to about 1 mm. Hg which is so slight that the person will become unconscious in only a few seconds and will die within another minute or so. Were it not for the carbon dioxide and water vapor, it would be possible at 50,000 feet to supply far more oxygen to the blood.

Respiratory compensation for oxygen deficiency—acclimatization. The chemoreceptor mechanism described in the previous chapter increases alveolar ventilation when a person develops oxygen deficiency at high altitudes. This mechanism unfortunately is not very powerful, for it normally can increase alveolar ventilation only about 60 per cent. Yet even this amount allows the aviator to ascend several thousand feet higher than he could otherwise.

When a person is exposed to high altitudes for several days at a time, his chemoreceptor mechanism, for reasons not well understood, becomes progressively more active until his alveolar ventilation increases to as much as five times normal. This slow process of developing more and more ventilation is called *acclimatization* to high altitudes.

Another factor that favors acclimatization is an increase in the number of red blood cells, for oxygen deficiency causes rapid production of these cells by the bone marrow, as explained in Chapter 9. Unfortunately, the formation of many new cells is a slow process, requiring several weeks to help acclimate a person to high altitudes. Persons who live at high altitudes all the time often develop red blood cell counts as high as 7 to 8 million per cu. mm., which is 50 per cent or more greater than normal.

Acclimatization to high altitudes is not of special importance in aviation because the aviator

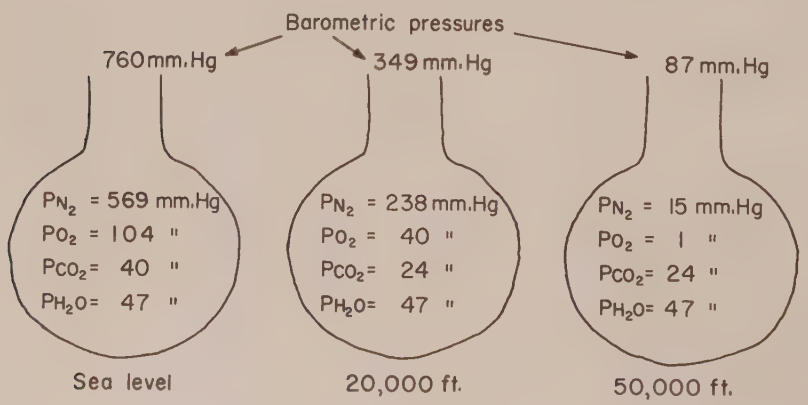


Figure 175. Effects of low barometric pressures at different altitudes on alveolar gas concentrations.

rarely remains aloft long enough for acclimatization to occur. It is far more important to the mountain climber, who must become slowly acclimated if he is to succeed in ascending to the top of the highest mountains. This explains why ascension is normally accomplished in slow stages, thereby allowing the body to become progressively acclimated. By using this procedure of acclimatization the conquerors of Mt. Everest were able to remove their masks for several minutes atop this highest mountain of the world, though this would have caused coma in a normal person in less than a minute.

Oxygen Breathing at High Altitudes

A person can ascend to far higher altitudes when he breathes pure oxygen than when he breathes air, because oxygen then occupies the space in the alveoli normally occupied by nitrogen, in addition to the usual oxygen. This allows the alveolar partial pressure of oxygen to remain quite high even when the barometric pressure falls to a low value. Observing Figure 175 once again, it will be noted that at 20,000 feet the combined pressure of oxygen and nitrogen is 278 mm. Hg. If the nitrogen were replaced with oxygen, the oxygen pressure would also be 278 mm. Hg, but because of the presence of the nitrogen the oxygen pressure is only 40 mm. Hg, which is low enough to cause considerable desaturation of the blood. When breathing pure oxygen at 20,000 feet the arterial hemoglobin is 100 per cent saturated with oxygen, though when breathing air the saturation is only 67 per cent.

Figure 176 presents graphically the percentage saturation of hemoglobin in the arterial blood at different altitudes when breathing air as opposed to breathing pure oxygen. It is evident from this figure that essentially no desaturation occurs when breathing pure oxygen below an altitude of approximately 33,000 feet. Above that point, however, the barometric pressure becomes so low that even with nitrogen eliminated from the alveoli, the pressure of oxygen is still not sufficient to keep the blood saturated. The lowest level of arterial oxygen saturation at which a non-acclimated person can remain alive is about 50 per cent. Therefore, the ceiling for a person breathing pure oxygen is approximately 47,000 feet, in comparison with a ceiling of 23,000 feet for one breathing air.

Pressurized Cabins

Obviously, all the problems of low barometric pressure at high altitudes can be overcome by

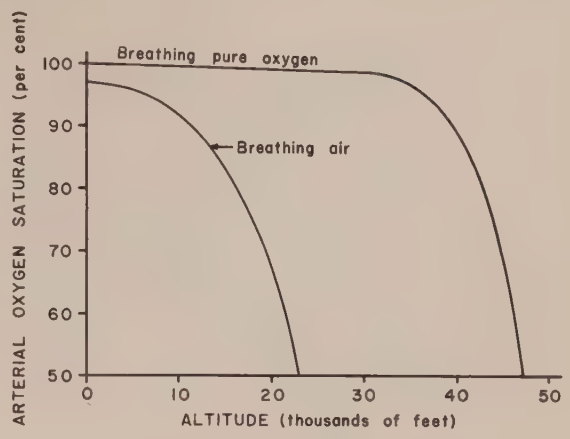


Figure 176. Relative effects of breathing air or pure oxygen on the saturation of arterial hemoglobin with oxygen at different altitudes.

pressurizing the airplane. Usually the air inside the cabin of passenger planes is pressurized to maintain approximately the same pressure as that at 5000 feet.

Explosive decompression. One of the major problems of pressurizing high altitude equipment is the possibility that the chamber might explode. Experiments have shown, fortunately, that sudden decompression does not cause any significant damage to the body because of the decompression itself. The danger lies in exposure to the low partial pressures of oxygen in the rarefied atmosphere. Referring back to Figure 174, one sees that at altitudes above 30,000 feet the person has only 30 to 60 seconds of consciousness, during which time he must institute appropriate life-saving measures, such as putting on an oxygen mask or parachuting to safety. If he parachutes without the aid of a special automatic opening device it is important to open the chute as soon as possible, because he is likely to lapse into coma and fall to earth without benefit of the parachute. On the other hand, if he opens the chute he is likely to lapse into coma and die because of the altitude. Fortunately, modern high-flying jets are equipped with special devices that eject the aviator out of the plane when need be and also automatically delay the opening of the chute until an appropriate altitude is reached.

ACCELERATION EFFECTS OF AVIATION

Airplanes move so rapidly and change their direction of motion so frequently that the body is

often subjected to severe physical stress caused by the sudden changes in motion. When the *velocity* of motion changes either positively or negatively the effect is called *linear acceleration.* On the other hand, when the *direction* of motion is changed the effect is called *angular acceleration.* In general, the forces induced by linear acceleration during normal flight of an airplane are not sufficient to cause major physiologic effects. But when an airplane turns, dives, or loops, the centrifugal forces caused by angular acceleration are frequently sufficient to promote serious derangements of bodily function.

Angular Acceleration

Positive g. Figure 177 illustrates an airplane going into a dive and then pulling out. While the airplane is flying level, the downward force of the aviator against his seat is exactly equal to his weight. However, as he begins to come out of the dive, he is pushed against the seat with much greater force than his weight because of centrifugal action. At the lowest point of the dive the force may be as much as six times that which would have been caused by normal gravitational pull. This effect is called *positive angular acceleration,* and the person is said to be under the influence of +6 g force, or, in other words, a force six times that of gravity.

Negative g. At the beginning of the dive in Figure 177 the airplane changes from level flight to a downward direction, which throws the pilot upward against his seat belt. He then is not exerting any force at all against his seat but instead is being held down by a force equal to three times his weight. This effect is called *negative angular acceleration,* and the amount of force exerted is said to be −3 g.

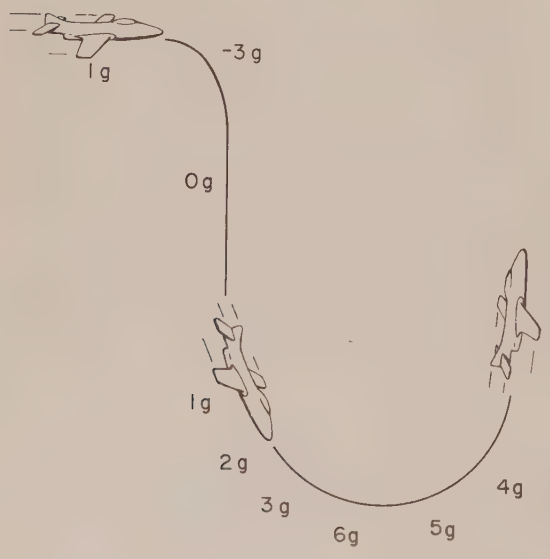


Figure 177. Positive and negative g during a dive and pullout of an airplane.

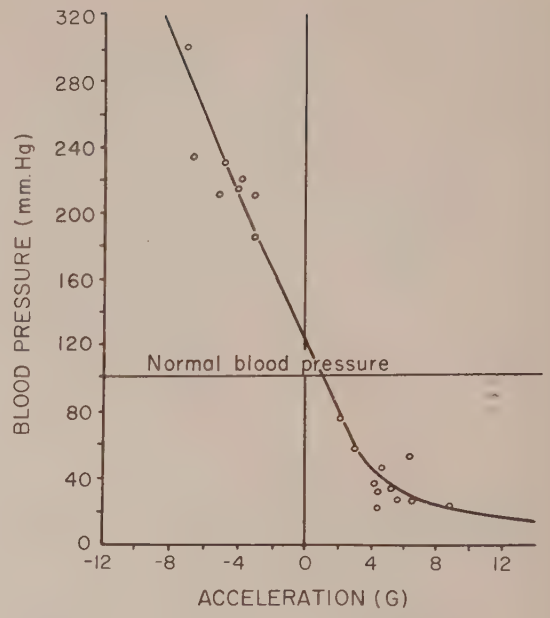


Figure 178. Effects of positive and negative g on the arterial blood pressure. (From Armstrong: *Principles and Practice of Aviation Medicine.* The Williams & Wilkins Company.)

Effect of angular acceleration on the circulatory system. The most important effects of angular acceleration on the body occur in the circulation. Positive g causes the blood in the vascular system to be centrifuged toward the lower part of the body. The veins of the abdomen and legs distend greatly, storing far more blood than usual, so that little or none of it returns uphill to the heart. The cardiac output falls to very low values or even to zero. The arterial pressure falls, as depicted in Figure 178, and the person lapses into coma. Negative g causes the opposite effects; so much blood is centrifuged into the upper part of the body that the cardiac output rises and the arterial pressure rises. Occasionally the person develops such high pressures in his cerebral vessels that brain edema results, and, rarely, vessels even rupture, causing serious mental impairment.

Figure 179 depicts graphically the time required for blackout (coma) to ensue at various degrees of positive acceleration. A person can usually withstand +4 g almost indefinitely; he might develop a certain amount of dizziness, but, in general, his circulatory control systems can maintain sufficient cardiac output to prevent blackout. However, at acceleration values above

+4 g, blackout usually occurs quite rapidly—at 5 g in about 8 seconds, at 12 g in about 3 seconds, and at 20 g in about 2½ seconds. Obviously, the amount of time that an aviator can remain conscious at high rates of acceleration determines how rapidly he can come out of a dive, and it also limits the sharpness of turns that he can make. Most airplanes are designed to withstand 20 or more positive g. Therefore, the maneuverability of an airplane is dependent more on what the aviator himself can stand than upon the strength of the plane.

ANTI-BLACKOUT MEASURES. Application of pressure to the outside of the legs and lower abdomen can prevent pooling of blood during positive acceleration. One means of doing this is for the aviator himself to tighten his abdominal muscles as tightly as possible during the pullout from a dive. This allows him to withstand a little more acceleration than he otherwise could and allows him several seconds extra exposure before blacking out. Different types of "anti-g" suits for accomplishing the same effect have also been devised. One of these utilizes air bags applied to the abdomen and legs. The bags are connected by a tube to the manifold of the engine, and when the airplane pulls out of the dive increased pressure in the engine applies positive pressure to the legs and abdomen. In this way the aviator is not uncomfortably subjected to pressure all the time but only when it is needed.

Because of the speed of the modern jet airplane, the tremendous acceleratory forces that can develop make it impossible for an aviator to turn sharply without blacking out. One of the solutions to this problem will probably be for the aviator to lie prone while flying the plane, for the body can withstand +15 g in the horizontal position in comparison with only +4 g in the sitting position.

Effect of positive acceleration on the skeleton. Also illustrated in Figure 179 is the effect of positive acceleration in causing vertebral fracture. This graph shows that with the aviator in the sitting position, positive acceleration greater than 20 g is likely to rupture a vertebra because of the intense force of the upper body pressing downward. Only a split second of the intense force can cause the fracture, for the stress that a bone can withstand does not depend on how long the stress is applied, but on how much stress is applied.

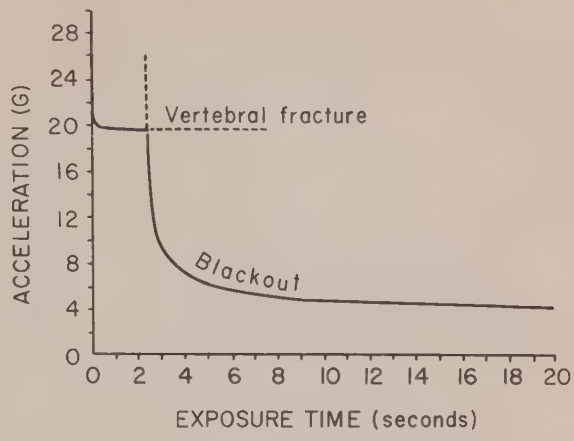


Figure 179. Effect of exposure time to different degrees of positive acceleration in producing blackout or vertebral fracture. (From Armstrong: *Principles and Practice of Aviation Medicine.* The Williams & Wilkins Company.)

Linear Acceleration

An aviator is occasionally subjected to linear acceleration when being catapulted into the air, and to a lesser extent when his plane takes off with JATO (jet-assisted takeoff). These effects are usually unimportant, though, if the aviator observes the precaution of sitting solidly back against his seat so that the airplane will not come forward suddenly and strike his head or other parts of his body.

Deceleration. Deceleration is negative linear acceleration, or, in other words, it is a progressive slowing of linear motion. In general, deceleration occurs so slowly in normal aviation that it has no effect on the body, but in accidents and on landing from a parachute descent the body is often decelerated so rapidly that serious damage results.

DECELERATORY FORCES IN ACCIDENTS. Figure 180 shows the principles and danger of deceleration in accidents. At the top a car is striking a concrete wall, and at the bottom it is striking a haystack. On striking the wall, deceleration occurs in the short stopping distance of less than an inch, whereas on striking a haystack, deceleration occurs over a distance of 10 feet. The ratio of the stopping distances is 120. Therefore, the force of impact against the haystack is 120 times less than that against the concrete wall. Obviously, because of the short stopping distance in the first instance, the person's body will be tremendously torn, while in the second instance it is likely to be unharmed.

Parachute descent. When a person jumps from an airplane, gravity causes him to begin falling toward earth. Figure 181 illustrates the velocity of fall at different distances (if the person has not opened his parachute). By the time he has fallen 1500 feet his velocity will have reached approximately 175 feet per second. At this speed, the air resistance exactly opposes the tendency for

$$\text{Force of impact} = \frac{\text{Energy}}{\text{Distance}}$$

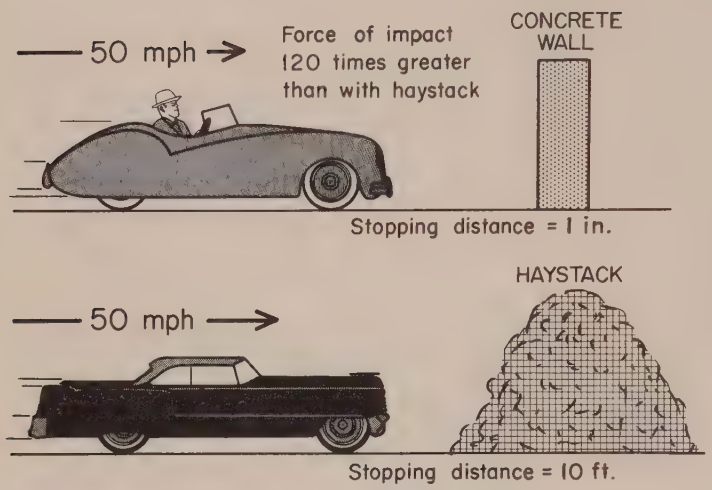


Figure 180. Effect of different distances of deceleration on the force of impact.

gravity to increase his velocity even more. Therefore, this velocity of fall, 175 feet per second, is the maximum that will be attained, and it is called the *terminal velocity.* (At very high altitudes the terminal velocity is considerably greater than 175 feet per second because of the rarefied atmosphere, but it will slow down to 175 feet per second when the person falls into the heavier atmosphere.)

One of the deceleratory forces caused by parachute descent is the *opening shock* that occurs when the parachute opens. The force of the shock is roughly proportional to the square of the velocity of fall at the time of opening. Parachutes are designed so that even after the person has reached the terminal velocity of fall the shock will not be sufficient to harm the body.

On landing, the parachutist approaches earth at a velocity of about 15 miles per hour. This is equivalent to jumping without a parachute from a wall approximately 8 feet high. Even this velocity can be dangerous if the parachutist is not ready to cope with the landing jolt. If he lands stiff-legged, he will almost certainly suffer a fracture of a leg, the pelvis, or a vertebra. On the other hand, if he does not tense his muscles when landing, he will pancake on the ground and likely suffer other injuries. Therefore, it is important that he land with his legs flexed but tense.

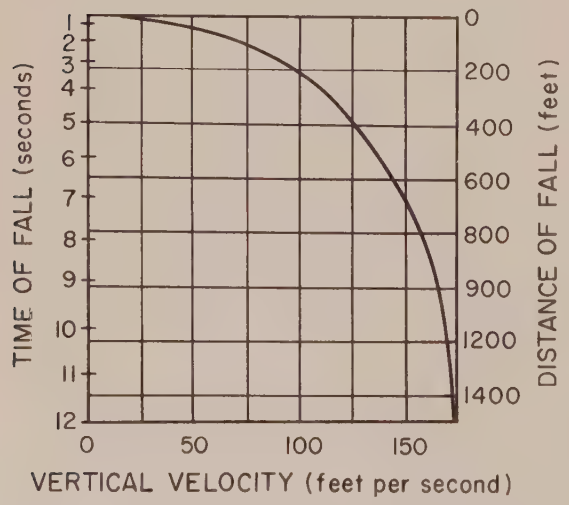


Figure 181. Effect of distance of fall on the velocity attained. (From Armstrong: *Principles and Practice of Aviation Medicine.* The Williams & Wilkins Company.)

SPACE PHYSIOLOGY

The physiologic principles of traveling in spaceships are much the same as those applicable to aviation physiology, except that the importance of some of the factors is intensified. The special problems in space physiology include: (a) weightlessness, (b) intense linear acceleration forces, (c) supply of oxygen and other nutrients, and (d) special environmental hazards.

Weightlessness. Prior to the advent of the spaceship, persons had experienced the phenomenon of weightlessness for not more than a few

seconds at a time, this state having been created in the cabins of airplanes undergoing a trajectory path, upward and over a hump, calculated for exact course and velocity to provide zero g forces for a period of 10 to 15 seconds. Yet it was known that once a spaceship was in orbit or was traveling from planet to planet, the spaceship itself would be subjected to exactly the same gravitational forces as the spaceman inside the ship. Therefore, the person would experience the phenomenon of prolonged weightlessness; that is, there would be no force pulling him in any direction toward any part of the spaceship.

Prior to the first space flight, many physiologists were concerned about how a human being would react to weightlessness of long duration. Fortunately, however, now that man has experienced this phenomenon, it has been proved to have no significant effect on the physiology of the body. However, its effects necessitate the use of a few operational devices for flying the spaceship and for other activities within the spaceship. First, since the spaceman "floats" inside the spaceship, he must be either strapped in place or have appropriate hand holds. Second, his food must be in special containers that have closed tops, because the food will literally float off any plate and fluids will float out of any glass. Ordinarily, foods are prepared in tubes of the toothpaste type, and the food is simply squeezed into the mouth. Likewise, water is sucked from a tube. Excreta must also be forced into a container rather than left to float freely in the spaceship. Other than for these simple problems, the state of weightlessness is actually very peaceful and pleasant, which is what one might suspect from his own experience of floating in water.

Linear acceleration of a spaceship. At takeoff, a high velocity is reached in only a few minutes while putting the spaceship into orbit. This is achieved by using rocket propulsion from one or more boosters. Figure 182 shows the intensity of acceleration at different times during the five minutes required to put a spaceship into orbit, illustrating that linear acceleration increases to 2 g when the first booster begins to fire. As that booster becomes lighter and lighter because of decreasing fuel, the acceleration increases to as much as 9 to 10 g. Then after the first booster has dropped, the second booster goes through a similar pattern of acceleratory changes. Also, the final stage, which carries the space capsule, accelerates for another minute or so before its rocket propulsion completely ceases. Upon cessation of all propulsion, the spaceman enters the state of weightlessness, which is designated in the figure by zero g acceleration.

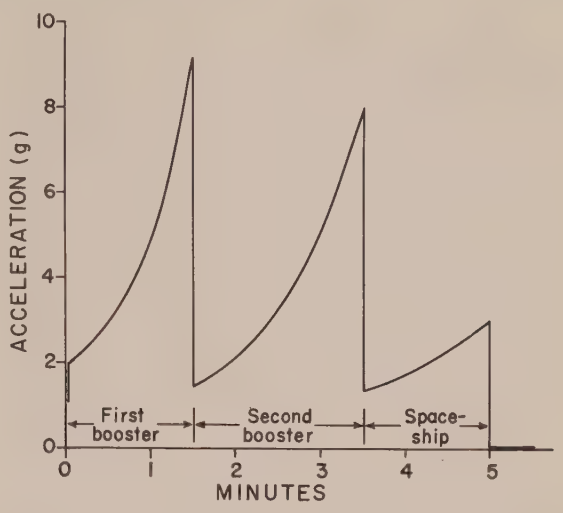


Figure 182. Acceleratory forces during takeoff of a spaceship.

In the sitting position the spaceman cannot withstand more than about 4 g, but in the reclining position, the position assumed in a reclining easy chair, the spaceman can withstand 12 to 15 g, which is greater than any of the forces illustrated in Figure 182. Therefore, this is the position that the spaceman assumes during takeoff.

DECELERATION. One of the greatest problems in all of space physiology is the effect of slowing the spaceship as it reenters the atmosphere. The tremendous amount of kinetic energy stored in the moving spaceship must be dissipated as the spaceship decelerates. This energy becomes heat and obviously increases the temperature inside the space capsule to very high levels. Therefore, special precautions are taken to provide a heat dissipation shield at the front of the spaceship which absorbs most of this heat rather than allowing it to be transmitted to the cabin of the ship. A spaceship traveling at a speed of Mach 100, a speed about three times that required for orbiting the earth, but which might well be used in interplanetary space flights, requires a distance of about 10,000 miles for safe deceleration. Any more rapid deceleration than this would create more g forces than the spaceman could stand.

Replenishment of oxygen and other nutrients. It is a simple matter to provide all the oxygen and nutrients needed by the spaceman if he is to remain in space for not more than a few days to a few weeks. However, interplanetary travel will require that a spaceman remain in space for many months or even years, for which reason it becomes physically impossible to carry sufficient oxygen and other nutrients within the confines allowable

in the spaceship. Therefore, many different procedures have been attempted to create a complete *life cycle* within the spaceship that will continually replenish the spaceship's oxygen and also supply adequate nutrients to the spaceman. For instance, algae are capable of living on human excreta, utilizing both carbon dioxide and fecal excreta to form oxygen, carbohydrates, proteins, and fats. Theoretically, algae can be used as food and the oxygen can be breathed. Unfortunately, though, the amount of algae necessary to provide this complete life cycle for spacemen is far greater than space limitations in the spaceship will allow. Furthermore, the foods developed from algae have proved to be neither palatable nor totally life sustaining. Yet future research might provide an adequate means for finally completing this life cycle.

Other procedures for completing the life cycle are based on chemical or electrochemical procedures for separating oxygen from carbon dioxide and for resynthesizing certain types of foodstuffs. Here again, the success has not been completely rewarding, so that today's space flights are still limited by the amount of oxygen and nutrients that can be carried from earth at the time of takeoff.

Radiation and other problems of space flights. Soon after space flights first started, it was discovered that there is ionizing radiation in space at altitudes above 600 miles, its intensity being far greater at some levels than at others. Fortunately, all manned space flights thus far have been at altitudes of 100 to 200 miles so that spacemen thus far have not been subjected to any hazard of ionizing radiation. Yet for interplanetary travel or for the establishment of space platforms at very high altitudes, sufficient radiation does exist to be a serious matter. For instance, even with the fastest spaceship, a person would receive approximately one fiftieth of a lethal dose of radiation as he passed through the radiation levels.

In addition to the natural radiation that occurs at very high altitudes, some of man's high altitude nuclear bomb explosions have added to the radiation hazard in space.

Other hazards at high altitudes include some of the same hazards of aviation but to a much severer degree, such as (1) exposure of the spaceman to ultraviolet radiation, which can cause severe sunburn or can blind him, (2) exposure to extreme heat when in the direct pathway of sunrays and to extreme cold when on the opposite side of a spaceship from the sunrays, and (3) exposure to very low barometric pressures in case of decompression of the spaceship. All of these problems require special engineering such as pressurized suits, protective filters for the eyes, and very intricate heat control systems for the space capsule.

DEEP SEA DIVING AND HIGH PRESSURE PHYSIOLOGY

The human body is occasionally exposed to extremely high barometric pressure such as when diving deep beneath the sea or when working in a compression chamber—for instance, in a tunnel beneath a river filled with compressed air to prevent cave-in. The very high barometric pressure increases the amount of gases that pass from the alveoli into the blood and become dissolved in all the body fluids. This extra dissolved gas, unfortunately, often leads to serious physiological disturbances.

Pressures at different depths of the sea. A person on the surface of the sea is exposed to the pressure of the air above the earth. This pressure is approximately 760 mm. Hg; or it is also stated to be 1 atmosphere pressure. The weight of 33 feet of water is equal to the weight of the atmosphere, so that at a depth of 33 feet beneath the sea

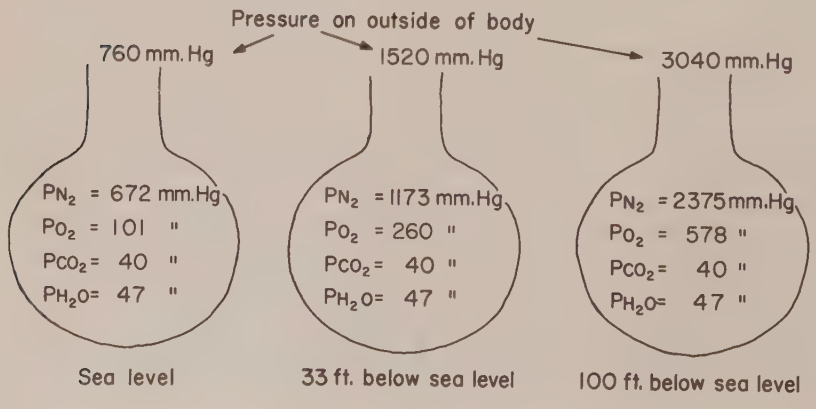


Figure 183. Effect of different sea depths on the gas pressures in the alveoli.

the pressure is double that at the surface, or 2 atmospheres of pressure. Thus, the pressure increases as one descends beneath the sea in accordance with the following table:

Feet	*Atmosphere(s) pressure*
0	1
33	2
67	3
100	4
200	7
300	10

Effect of High Gas Pressures on the Body

Figure 183 illustrates the pressures of each of the alveolar gases when a diver is breathing compressed air at 33 feet below sea level, and also at 100 feet below sea level. It will be noted that the pressures of water vapor and carbon dioxide remain the same regardless of how far below the surface the person descends. The reason for this is that water vapor pressure is directly dependent on the temperature of the body, as explained in Chapter 17, and the pressure of carbon dioxide is dependent on the rate of carbon dioxide release from the blood into the lungs, which continues to be the same regardless of depth. On the other hand, the pressures of nitrogen and oxygen increase almost directly in proportion to the depth below sea level to which the person descends.

Oxygen poisoning. Ordinarily a person breathing compressed air can descend to about 200 feet below the sea without any danger from excess oxygen absorption. However, at levels deeper than this, his tissues begin to be damaged very rapidly by the oxygen. The reason for this is the following: Hemoglobin normally releases oxygen to the tissues at pressures ranging between 20 and 45 mm. Hg. In this way the hemoglobin automatically "buffers" the oxygen in the tissues, maintaining a relatively constant oxygen pressure that allows a relatively constant rate of cellular oxidation. When there are extreme oxygen pressures in the lungs, however, large quantities of oxygen dissolve in the fluids of the blood, and the oxygen transported in this manner is released to the tissues at pressures that are much greater than 45 mm. Hg, supplying far more than normal amounts of oxygen to the cells. This causes deranged cellular metabolism, often damaging the cells themselves or leading to abnormal cellular function. The most debilitating effect of oxygen poisoning occurs in the brain, which is usually manifested by nervous twitchings, convulsions, and coma. Obviously, the person often loses control of himself, which might make him descend too deeply or fail to operate his breathing apparatus correctly, which will lead to his death.

The dangers of oxygen poisoning can be eliminated by supplying a progressively less and less percentage of oxygen in the diver's breathing mixture as he goes to greater depths.

Nitrogen poisoning. High pressures of nitrogen can also seriously affect a diver's mental functions even though it does not enter into any metabolic reactions in the body. When it dissolves in the body fluids in high concentrations, it exerts an anesthetic effect on the central nervous system. Ordinarily, a person can descend to approximately 200 feet below sea level before this anesthetic effect becomes serious, which is about the same safe depth as that for the prevention of oxygen poisoning. However, at depths slightly lower than this the diver will actually fall asleep because of the increasing amounts of nitrogen dissolved in his fluids. Also, in the early stages of nitrogen narcosis, the person frequently develops a sense of extreme exhilaration associated with seriously depressed mental acuity, a condition called "raptures of the depths." This state is comparable to severe drunkenness from alcohol, and it can cause the diver to descend much deeper than he should or to perform other dangerous acts that will lead to his death.

To avoid the dangers of nitrogen poisoning, the nitrogen is frequently replaced with helium in the breathing mixture. Helium does not cause an anesthetic effect, and it has an additional advantage of diffusing out of the body fluids more rapidly than nitrogen would when the diver ascends.

The problem of carbon dioxide washout from the diving helmet. If high concentrations of carbon dioxide are allowed to collect in a diver's helmet, he will soon develop respiratory acidosis and perhaps even coma. Therefore, carbon dioxide must be washed out of the helmet rapidly enough so that its concentration will never rise to any significant level. To do this, the same volume of gas must flow through the helmet at all times. When the diver descends to 200 feet below the sea, his air must be compressed seven times to withstand the pressure. This also reduces the volume seven times, and requires that the compressor pump seven times as much air for carbon dioxide washout as is required at sea level. It can be seen, therefore, that the rate at which air is supplied to the diver must be increased directly in proportion to the increase in pressure at which he is working. Indeed, it is this factor of carbon dioxide washout that determines how much air must be pumped to the diver rather than the amount of oxygen that he needs.

SCUBA DIVING. The term SCUBA means "self-contained underwater breathing apparatus." The major factor that limits the time a person can remain under water using a SCUBA system is the problem of carbon dioxide washout from the alveoli. The deeper the diver goes, the less becomes the *volume* of each unit mass of gas liberated from his tank. Therefore, the greater must be the *mass* of gas that flows out of his tank to maintain carbon dioxide washout from his lungs. Thus, a tank of compressed air will last only one seventh as long at 200 feet as at a few feet below the surface of the sea. For this reason, a SCUBA diver can descend to a 200-foot depth for only 15 to 20 minutes at a time. Other than for this limitation, the physical principles of SCUBA diving are essentially the same as those of helmet diving.

Decompression Sickness

Solution of nitrogen in the body fluids. In addition to the anesthetic effect of nitrogen, this gas can also cause very serious damage because of bubble formation in the body fluids when the diver returns toward the surface. The amount of nitrogen normally dissolved in the entire body when a person is exposed to normal atmospheric pressure is about 1 liter. If he remains 100 feet below the sea for several hours, that is, at 4 atmospheres pressure, the amount of dissolved nitrogen increases to approximately 4 times as much, and at 7 atmospheres pressure to 7 times as much.

Bubble formation during decompression. Figure 184 illustrates to the left the gases dissolved in the body fluids when a diver is breathing air at a pressure of 5000 mm. Hg, which corresponds to a depth below the sea of approximately 200 feet. The pressure on the outside of the body is 5000 mm. Hg, and the pressure of all the dissolved gases in the body fluids is only 4045 mm. Hg. Therefore, the pressure in this instance is great enough on the outside of the body to keep the dissolved gases from ever becoming bubbles. To the right in this same figure, however, the person has suddenly been brought to the surface of the sea where the pressure on the outside of his body and in his lungs is only 760 mm. Hg. Immediately, the gases begin to expand inside the fluids because the pressure on the outside of the body is no longer sufficient to keep so much gas dissolved. The expanding gases form bubbles throughout the body —in the cells, in the blood, in the spinal fluid, and almost everywhere else.

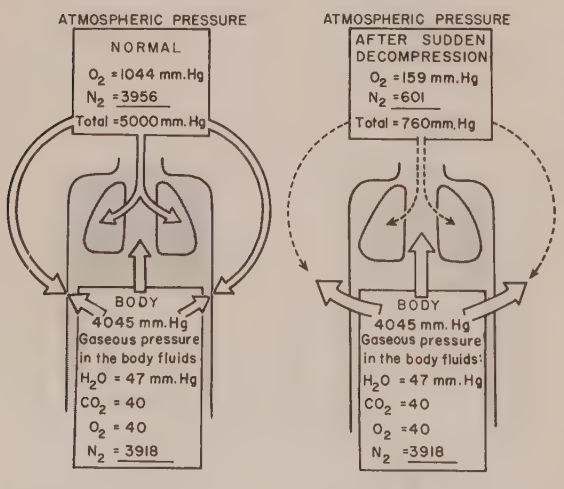


Figure 184. The dynamics of bubble formation during sudden decompression.

Effect of bubble formation on the body. The formation of bubbles in the body is known by many different names: decompression sickness, the bends, diver's paralysis, caisson disease, and others. The most distressing effects occur when bubbles develop in the central nervous system. They sometimes occur inside neuronal cells, but mostly in extraneuronal tissues of the brain and spinal cord, causing mechanical rupture of fiber pathways. The damage can lead to serious mental disorders and very frequently to permanent paralysis because of ruptured fiber pathways to the muscles.

Decompression sickness usually causes severe pain, which may result from bubble damage in the central nervous system or in peripheral nerves. Sometimes the pain is caused by bubble formation in pain-sensitive tissues such as the joints and bones, or, finally, it can result from sudden distention of the gastrointestinal gases, causing severe bloating of the gut.

In short, many different symptoms are produced by bubble formation and gas expansion in the body, though the most distressing and permanent of these result from damage to the nervous system.

Prevention of decompression sickness. A total of approximately 2 liters of bubbles can usually develop in the body fluids without dangerous consequences, but more than this almost always causes symptoms. These can be prevented, however, by allowing the diver to come to the surface very slowly or by decompressing him in a decompression chamber. In either event, the pressure around his body must be decreased so gradually that the excess gases dissolved in his fluids can leave through his lungs before the quantity of bubbles rises above 2 liters.

Decompression sickness in aviation. Rarely, an aviator ascends in an airplane so rapidly that

he, too, forms more than 2 liters of bubbles and develops decompression sickness. It may be difficult to understand how so great a volume of bubbles can form when only 1 liter of nitrogen is normally dissolved in the body fluids. However, at high altitudes the liter of nitrogen expands in proportion to the decreased pressure. For instance, at 50,000 feet it becomes approximately 8 liters which is certainly adequate to elicit decompression sickness.

REFERENCES

Armstrong, H. G.: *Principles and Practices of Aviation Medicine.* 3rd Ed. Baltimore, The Williams & Wilkins Company, 1952.

Bibliography of Space Medicine. Washington, National Library of Medicine, Reference Division, 1958, and *U. S. Armed Forces Med. J.* 10:172, 1959.

Fulton, J. F.: *Decompression Sickness.* Philadelphia, W. B. Saunders Company, 1951.

Korner, P. I.: Circulatory adaptations in hypoxia. *Physiol. Rev.* 39:687, 1959.

Roxburgh, H. L., and Erasting, J.: The physiology of pressure suits. *J. Aviat. Med.* 28:260, 1957.

Space travel: a symposium. *J. Aviat. Med.* 28:479, 1957.

Stickney, J. C., and Van Liere, E. J.: Acclimatization to low oxygen tension. *Physiol. Rev.* 33:13, 1953.

Submarine Medicine Practice, Department of the Navy. Washington, U. S. Government Printing Office, 1956.

U. S. Navy Diving Manual. Washington, U. S. Government Printing Office, 1958.

THE NERVOUS SYSTEM AND MUSCLE

THE NERVE, AND MEMBRANE POTENTIALS

20

The function of a nerve fiber is to conduct signals from one part of the body to another. Nerve fibers are long thread-like extensions of nerve cells, called *neurons,* whose cell bodies are located mainly in the brain and spinal cord, though a few are in ganglia outside the central nervous system.

Nerve fibers extend from one part of the central nervous system to another part or into the peripheral nerves. The purpose of the present chapter is to explain the mechanism by which signals are transmitted over nerve fibers, but to do this it is necessary to explain first how electrical potentials are generated at the nerve fiber membrane. We shall see in later chapters that the mechanism for transmission of impulses along nerve fibers also applies to transmission of impulses in muscle fibers and over the surfaces of nerve cell bodies in the central nervous system.

Physiologic Anatomy of the Nerve Fiber

Figure 185A illustrates a neuron with its long fiber extension called an *axon,* and Figure 185B shows a cross-section of the axon itself. The function of the cell body will be discussed in Chapter 22. The present chapter is concerned with the conduction of impulses along the axon. Referring again to Figure 185B, one sees that the axon contains in its center a gelled substance called *axoplasm,* and this is surrounded by a *membrane* that

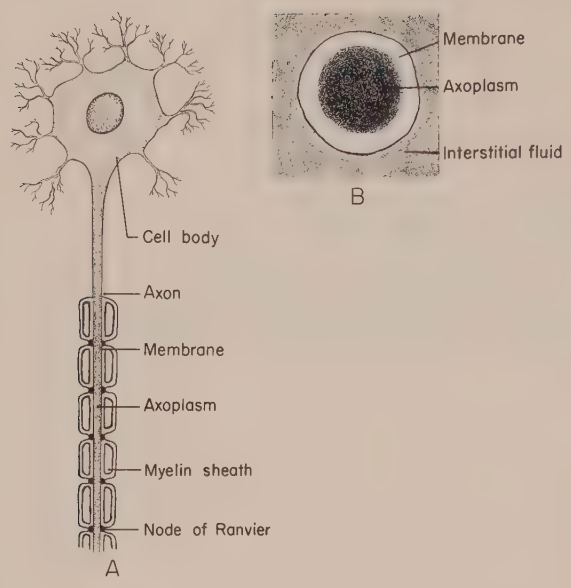


Figure 185. (A) A nerve cell with its thread-like axon, showing the nerve membrane, the axoplasm, and the myelin sheath. (B) Cross-section of an axon.

separates the axoplasm from the interstitial fluid. This membrane has the same functions as any other cellular membrane, except that it is specifically adapted for transmission of electrochemical impulses.

As illustrated in Figure 185A, some nerve fibers are surrounded on their outsides by an additional layer, the *myelin sheath,* which is discontinuous at the *nodes of Ranvier.* The myelin sheath is an insulator, the function of which is explained later in the chapter.

MEMBRANE POTENTIALS

The first essential for conduction of a nerve impulse is the establishment of an electric potential across the axon membrane. This potential is called a *membrane potential.* It results from concentration differences of certain ions between the two sides of the membrane.

Development of concentration differences. The membrane of the axon, like other cellular membranes, actively transports some ions from the interstitial fluid into the axoplasm and others from the axoplasm back into the interstitial fluid. For nerve conduction, the most important of these transfer processes is the active transport of sodium from the axoplasm into the interstitial fluid, which is often referred to as the *sodium pump.* This, by making the quantity of sodium inside the axoplasm very little, causes a membrane potential to develop as follows:

In Figure 186, the concentration of sodium inside the nerve fiber is shown to be approximately 10 milliequivalents, in contrast with its concentration of 137 milliequivalents in the interstitial fluid. On the other hand, the potassium concentration inside the nerve fiber is approximately 35 times its concentration in the interstitial fluid. However, potassium ions can diffuse in either direction through the membrane with relative freedom. The high concentration of potassium on the inside of the axon is mainly a secondary effect of the sodium pump as follows: When sodium ions, which are positive ions, are pumped to the outside a great void of positive ions is left inside the membrane. As a result, the inside becomes strongly electronegative, and, since potassium ions, which are also positive ions, are freely diffusible through the membrane, the electronegativity pulls a large quantity of these ions to the inside, thus partially filling the deficit left by the expulsed sodium ions.

The electronegative potential created inside the fiber as a result of the sodium pump is the membrane potential. In Figure 187A the membrane potential is illustrated by a series of positive and negative charges, positivity outside the membrane and negativity inside. This is the normal resting state of the nerve fiber, and the actual voltage of the membrane potential normally is about 0.085 volt or 85 millivolts.

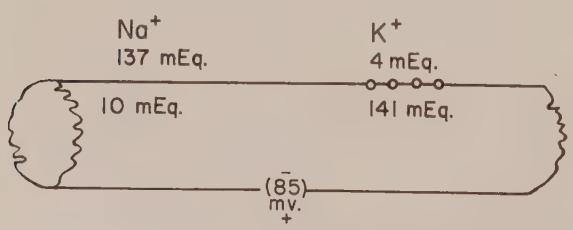


Figure 186. Concentration gradients of sodium and potassium at the axon membrane, showing that the membrane in the resting state is permeable only to potassium.

Depolarization of the Membrane, and the Nerve Impulse

Figure 187B illustrates at its central-most point an area that has suddenly become so permeable that even sodium ions can now diffuse through the membrane with ease. As a result, sodium ions (which are positive) rush to the inside, causing the membrane to become suddenly positive inside

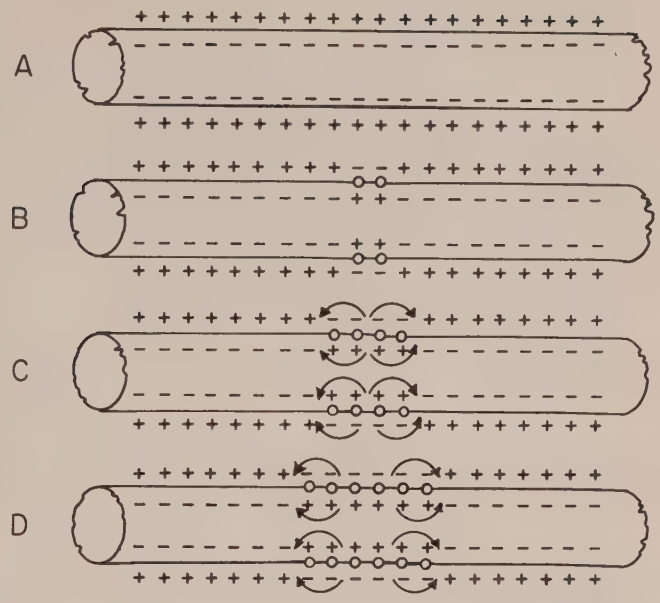


Figure 187. Transmission of the depolarization wave, the initial event in the nerve impulse.

and negative outside. This is the opposite to the usual resting state of the membrane, and it is called *depolarization* because the normal polarized state, with positivity on the outside and negativity on the inside, no longer exists.

The depolarization wave or nerve impulse. In Figures 187C and D, the area of depolarization in the middle of the fiber has extended in both directions, and the area of increased permeability has also extended. The cause of this extension is the flow of electrical current from the original depolarized area to adjacent areas. This passage of electrical current, for reasons not yet completely understood, makes the adjacent areas permeable also to sodium. The sodium ions flow through the membrane at these new areas, depolarizing these areas and causing still more electrical current to spread along the fiber, thereby extending the area of permeability still further. Again more sodium flows through the membrane, more electrical current travels along the fiber, and the cycle repeats itself over and over again. This spread of increased permeability and electrical current along the membrane is called a *depolarization wave* or a *nerve impulse.*

It is obvious from the preceding discussion that once a single point anywhere on the nerve becomes depolarized a nerve impulse travels away from that point in each direction, and each impulse keeps on traveling until it comes to both ends of the fiber. In other words, a nerve fiber can conduct an impulse either toward the nerve cell or away from it.

Repolarization of the Nerve Membrane

Immediately after a depolarization wave has traveled along a nerve fiber, the inside of the fiber is positively charged because of the large number of sodium ions that have diffused to the inside. This state of events causes the membrane to become impermeable to sodium ions again, though the exact mechanism of this is not understood. Yet potassium can still diffuse through the pores with relative ease, and because of the high concentration of potassium on the inside many potassium ions diffuse outward carrying positive charges with them. This once again creates electronegativity inside the membrane and positivity outside, a process called *repolarization.*

Repolarization usually begins at the same point in the fiber where depolarization had originally begun, and it spreads along the fiber in the manner illustrated in Figure 188. Repolarization occurs a few ten-thousandths of a second after depolarization. That is, the whole cycle of depolar-

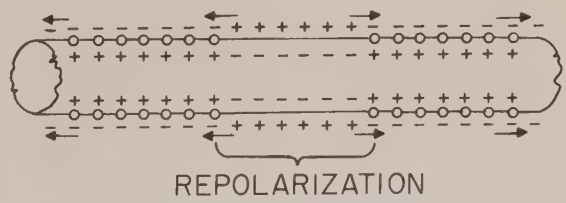


Figure 188. The repolarization process, the concluding event in the nerve impulse.

ization and repolarization takes place in a very minute fraction of a second, and the fiber is then ready to transmit a new impulse.

REFRACTORY PERIOD. When an impulse is traveling along a nerve fiber, the nerve fiber cannot transmit a second impulse until the fiber membrane has been repolarized. Even a strong electrical stimulus causes no effect, for which reason the fiber is said to be in a *refractory* state, and the time that the fiber remains refractory is called a *refractory period.* This varies from about ½500 second for large nerve fibers up to ½250 second for small fibers.

Reestablishment of Ionic Concentrations after Nerve Impulses Are Conducted

After the nerve fiber is repolarized the sodium ions that have leaked to the inside of the membrane and the potassium ions that have leaked to the outside must be returned to their original sides of the membrane. This is accomplished by the sodium pump which was discussed previously. That is, this mechanism pumps sodium ions to the outside, creating a deficit of positive charges inside the membrane. The resulting electronegativity inside the membrane then pulls potassium ions back to the inside. Thus, this process restores the ionic differences to their original levels so that additional nerve impulses can travel along the membrane.

However, it must be emphasized that even when the sodium pump fails to act, a hundred thousand or more impulses can still be transmitted over a single nerve fiber before transmission will cease. The reason for this is that only a few trillionths of a mole of sodium enter the fiber each time a single impulse is transmitted, so that it takes a hundred thousand or more impulses for enough sodium to enter the fiber to cause cessation of impulse transmission. It is evident, then, that the sodium pump is not necessary for repolarization of the membrane after each nerve impulse; this is accomplished by the diffusion of potassium outward through the membrane pores. Instead, the sodium pump simply plods along slowly, reestablishing ionic concentration differences across the membrane whenever a train of impulses tends to alter these.

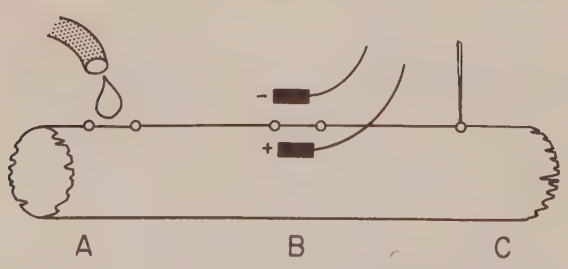


Figure 189. Different means of stimulating the nerve fiber.

Types of Stimuli That Excite the Nerve Fiber

Some of the types of stimuli that can initiate a nerve impulse are shown in Figure 189. At point A the fiber membrane is stimulated by a chemical substance such as an acid, a base, or even a strong electrolytic solution. These alter the membrane so that it becomes permeable to sodium, which initiates an impulse in each direction along the fiber. At point B the membrane is excited electrically. A negative charge is applied outside the membrane and a positive charge inside, which is opposite to the normal polarity of the resting membrane potential. When a reverse polarity such as this is applied, the membrane becomes highly permeable, and an impulse begins in each direction. Finally, at point C a needle is shown pricking the membrane. Trauma caused in this manner or by crushing, burning, or otherwise damaging the fiber can increase the permeability of the membrane and result in an impulse.

In the living body, nerve fibers normally are stimulated by both physical and chemical means. For example, pressure on certain of the nerve endings in the skin mechanically stretches these endings, in this way setting off impulses. Heat and cold affect other nerve endings to elicit impulses, and damage to the tissues, such as cutting the skin or stretching the tissues too much, can result in pain impulses.

In the central nervous system, impulses are transmitted from one neuron to another mainly by chemical means. The nerve ending of the first neuron secretes a chemical substance that in turn excites the second neuron. In this way impulses are sometimes passed through many hundred neurons before stopping.

In the laboratory, nerve fibers are usually stimulated electrically. A typical stimulator is illustrated in Figure 190. It emits electrical impulses of any desired voltage either singly or in rapid succession. Usually, a small probe having two wires at its end is placed on either side of a nerve trunk, and the electrical stimulus is applied so that electrical current flows through the fibers. As the current passes through the fiber membranes the permeability is altered, eliciting nerve impulses.

The all or none law. From the preceding discussion of the nerve impulse it is evident that if

a stimulus is strong enough to cause a nerve impulse at all, it causes the entire fiber to fire. In other words, a weak stimulus never causes a weak nerve impulse; either it is strong enough to stimulate the fiber or not. This is called the *all or none law.*

Minimal and maximal stimulus. Some of the fibers in a nerve trunk are more susceptible than others to stimulation. Therefore, as the voltage of an electrical stimulus is progressively increased from zero upward, a voltage will soon be reached at which a single nerve fiber will be stimulated. This voltage—that is, the least possible voltage that will cause even a single fiber to fire—is called the *minimal stimulus.* For instance, in Figure 191 the minimal stimulus of a particular nerve trunk is shown to be approximately 0.8 volt. As the voltage is increased still further, more and more nerve fibers become stimulated until finally a voltage is reached at which all the fibers in the trunk are stimulated. This is called the *maximal stimulus.* In Figure 191 the maximal stimulus of the particular nerve trunk being studied is 2.1 volts. All stimuli with voltages above the maximal stimulus always cause the same reaction, for all fibers are stimulated in each instance. In physiological experiments it is usually desirable to stimulate nerves with a stimulus voltage several times as great as the maximal stimulus because it is always important to know that all the nerve fibers are being stimulated. If one takes a chance on stimulating with a voltage less than maximal stimulus, he never knows precisely how many of the fibers in the trunk are being stimulated.

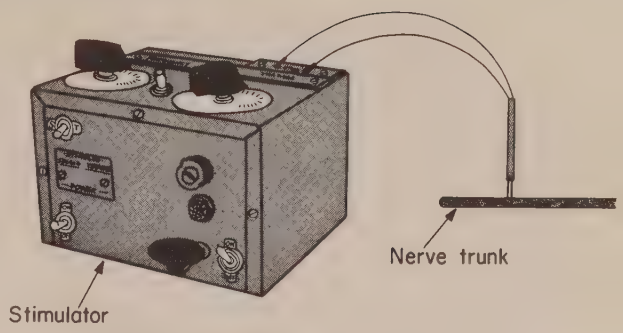


Figure 190. A laboratory stimulator.

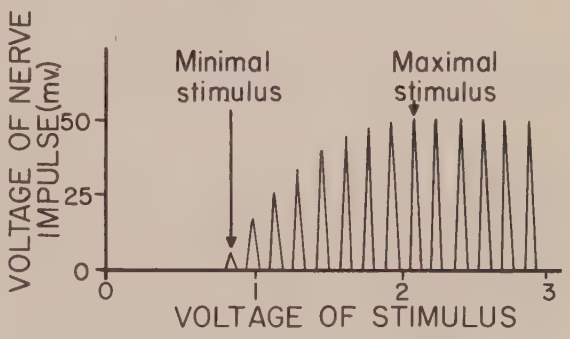


Figure 191. Minimal and maximal stimuli.

ACTION POTENTIALS

The changing electrical voltage at the nerve membrane when an impulse travels along the fiber is called an *action potential.* The method for recording potentials between the inside and outside of the nerve fiber is illustrated in Figure 192. This figure shows a minute glass pipet having a tip only one micron in diameter and filled with a strong solution of potassium chloride that conducts electricity from the tip. The tip pierces the nerve membrane to make contact with the fluid of the axoplasm. On the outside of the fiber is another electrode illustrated by the dark rectangle. These two electrodes are connected to a recording meter.

The changing potentials across the membrane are illustrated schematically in Figure 193A. This shows that during the original resting state, a potential of +85 millivolts is recorded outside

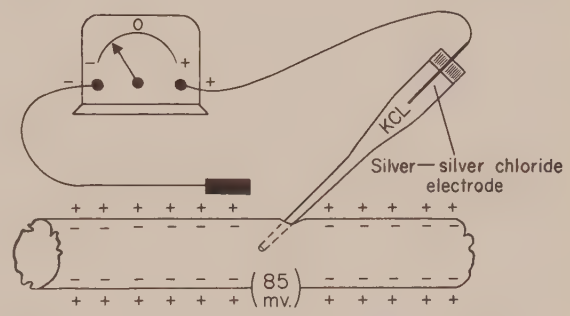


Figure 192. Measurement of membrane potentials and action potentials of a nerve fiber by means of a microelectrode.

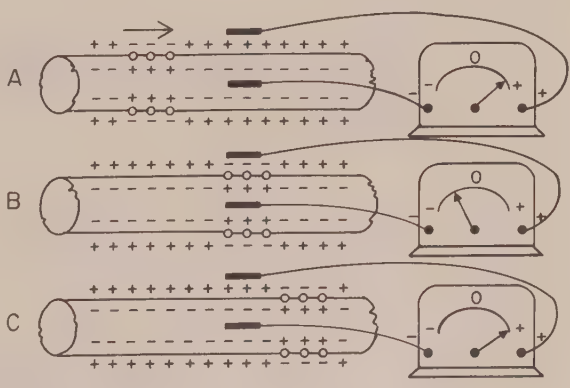


Figure 193. Principles of recording a monophasic action potential.

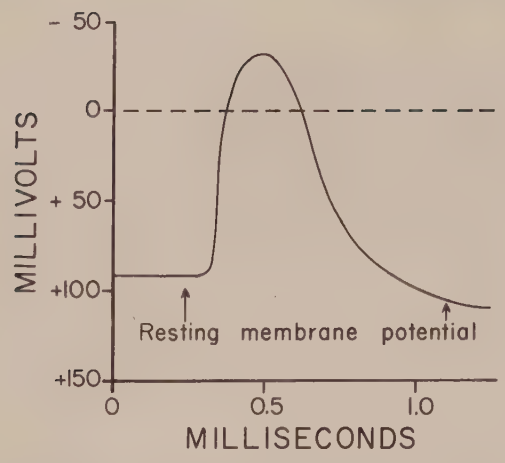


Figure 194. Graphic record of a monophasic action potential recorded from a large nerve fiber.

the fiber with respect to the inside. In Figure 193B a depolarization wave has traveled down the fiber until it is directly between the electrodes. At this instant, the rapid influx of positive sodium ions to the inside of the fiber reverses the potential, causing negativity outside the fiber and positivity inside as illustrated by the dial of the meter. Then, immediately after the impulse has passed the electrodes, the potential returns again to approximately the original resting potential because of diffusion of positive potassium ions to the outside.

Figure 194 illustrates a continuous recording of the potential changes as a nerve impulse passes the electrodes. This record is called the *monophasic action potential.* It begins with a normal resting membrane potential of +85 millivolts on the outside of the fiber and during the peak of the action potential reverses to become about −35 millivolts. Within approximately one-half millisecond, the normal membrane potential recovers, and the record returns almost to its original level. Actually, it "overshoots" a little, one or two millivolts, because the sodium pump immediately begins to pump positive sodium ions to the outside, which increases the potential outside slightly above 85 millivolts. This slight overshoot is called the *positive after potential.* It lasts for as long as 50 to 100 milliseconds, but by this time all the sodium that has entered the fiber during the action potential will have been removed from the fiber so that the original resting membrane potential is established again.

Use of the oscilloscope to record action potentials. Mechanical recording apparatus cannot function quickly enough to record the rapidly transient voltages of nerve action potentials. Therefore, a special instrument called the *oscilloscope,* which is similar to a television receiver, is normally used for this purpose. The principal components of the oscilloscope are shown in Figure 195. The basic part of this instrument is a cathode ray tube. An *electron gun* at the base of this tube projects toward the face of the tube a fine beam of electrons having a diameter of about one millimeter. The beam passes between four metal plates, two of which are placed on the two sides of the beam and the other two of which are above and below the beam. On the face of the cathode ray tube is a fluorescent material that glows brightly when the electron beam strikes. The beam in turn can be moved back and forth or up and down across the face of the tube by applying electrical potentials to the horizontal and vertical plates.

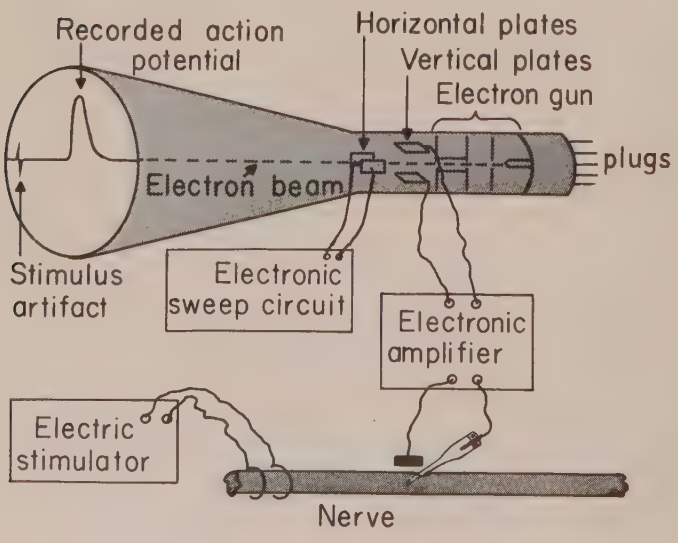


Figure 195. Diagram of an oscilloscope for recording action potentials from nerves.

An *electronic sweep circuit,* connected to the horizontal plates, causes the electron beam to move slowly from left to right across the face of the tube. When the beam has moved all the way across it jumps rapidly back to the left side of the face and begins moving across again, retracing the same pathway over and over again. The fluorescence along the path of the moving beam gives the appearance of a horizontal line on the face of the tube.

If, while the beam of electrons is moving across the tube, electrical potentials are applied to the vertical plates, the beam can be made to move up or down. In Figure 195 the beam of electrons deviates from the line slightly at the point called *stimulus artifact,* and then it deviates at the point called *recorded action potential.* The electrical voltages across the nerve membrane are responsible for these deviations. When the nerve is stimulated by an electrical stimulator, some of the electrical current from the stimulator spreads through the fluids of the nerve to the pickup electrodes and causes the stimulus artifact. At the same time, an action potential begins traveling down the nerve fiber toward the two pickup electrodes. Just as soon as the depolarization wave reaches the electrodes, the amplified potential applied to the vertical plates makes the electron beam move upward to record the action potential.

Most action potentials of nerves have a total duration of not more than a few ten-thousandths to a few thousandths of a second. The cathode ray oscilloscope fortunately can record any electrical potential that lasts at least a few millionths of a second. Therefore, it is quite capable of giving a true record of nerve action potentials.

TRANSMISSION OF SIGNALS OVER NERVE TRUNKS

Types of Fibers in the Nerve Bundle

Figure 196 shows a nerve bundle with fibers of different sizes and of two different types. The large black dots surrounded by white areas are *myelinated fibers;* the white area is a tube of insulator material called *myelin.* This material greatly increases the velocity of the impulses traveling along these fibers. In addition to the myelinated fibers are about twice as many small fibers without myelin sheaths. These are called *unmyelinated fibers.*

The large myelinated fibers control the rapid movements of the body and also transmit many of the sensory signals from the body to the brain.

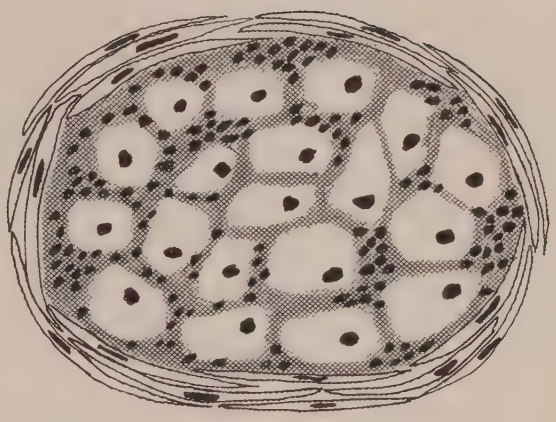


Figure 196. Cross-section of a nerve trunk, showing large, myelinated nerve fibers, small, unmyelinated nerve fibers, and the nerve sheath.

The unmyelinated fibers, which conduct impulses quite slowly, cannot cause rapid reactions. These control most of the subconscious activities of the body, such as the excitability of the heart, the contractions of the blood vessels, the gastrointestinal movements, and the emptying of the urinary bladder. Also, the small, unmyelinated fibers transmit those sensory signals that do not require immediate action, such as the aching type of pain sensations, crude touch sensations, and pressure sensations. Further descriptions of the different types of fibers and their functions will be presented in connection with the discussions of sensory functions of the nervous system in Chapter 23 and of motor functions in Chapter 24.

Function of the myelin sheath. Myelin, which surrounds all large nerve fibers, is a lipid substance that will not conduct electric current. Therefore, it acts as an insulator around the nerve fiber. As illustrated schematically in Figure 197, at approximately every millimeter along the length of the fiber the myelin is broken by a so-called *node*

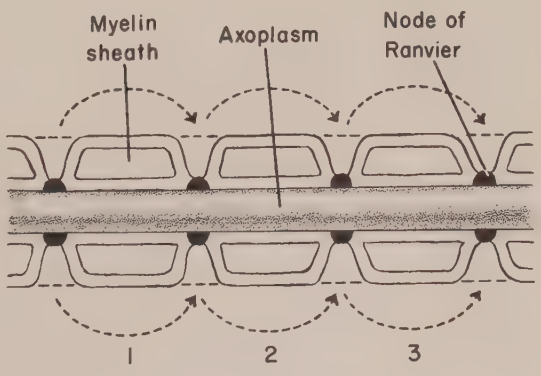


Figure 197. Saltatory conduction in a myelinated axon.

of Ranvier. At these nodes, typical membrane depolarization can occur, but beneath the myelin sheath, membrane depolarization does not take place. Instead, impulses are transmitted along a myelinated nerve by a process called *saltatory conduction* which may be explained as follows: Referring once again to Figure 197, let us assume that the first node of Ranvier becomes depolarized. This causes electrical current to spread around the outside of the myelin sheath all the way to the next node of Ranvier, causing it to become depolarized also. Current generated by this node causes the same effect on the next node; thus, the impulse "jumps" from node to node.

Saltatory conduction is believed to be valuable for two reasons: First, by causing the depolarization process to jump long intervals along the nerve fiber, it is believed to increase the velocity of conduction along the fiber. Second, and perhaps much more important, saltatory conduction prevents the polarization of large areas of the fiber and thereby prevents leakage of large amounts of sodium to the inside of the fiber. This conserves the energy required by the sodium pump to expel the sodium. Therefore, it is believed that the myelin sheath greatly decreases the amount of energy required by the nerve for impulse transmission.

Velocity of conduction in nerve fibers. The larger the nerve fiber and the thicker the myelin sheath, the more rapidly can the nerve conduct an impulse. The largest nerve fibers are about 20 microns in diameter, and the smallest are about 0.5 micron. The very large fibers conduct impulses at a velocity of 100 meters—about the length of a football field—in 1 second, while the very small fibers conduct impulses at a velocity of only 0.5 meter per second, or approximately the distance from the foot to the knee. All sizes of nerve fibers exist between these smallest and largest sizes, so that a wide spectrum of impulse velocities occurs in the different nerves.

Number of impulses that can be transmitted per second. The number of impulses that can be transmitted by any one fiber per second is determined by the refractory period of the fiber, and this depends to a great extent on the size of the fiber. Large fibers (15 to 20 microns in diameter) become repolarized in approximately $\frac{1}{2500}$ second. Therefore, a second nerve impulse can be transmitted $\frac{1}{2500}$ second after the first, or a total of up to 2500 impulses can be conducted each second. At the opposite extreme, the very smallest nerve fibers require as long as $\frac{1}{250}$ second to repolarize, which means that they can transmit no more than 250 impulses per second.

Transmission of Signals of Different Strengths by the Nerve Bundle

The nerve bundle has two means by which it can transmit signals of different strengths. These are (1) to transmit impulses simultaneously over varying numbers of nerve fibers, and (2) to transmit impulses in small or large numbers over the same fiber. These two means are called *spatial* and *temporal summation* respectively. As an example of spatial summation, if 100 nerve fibers are connected between the spinal cord and a certain foot muscle, stimulation of one of these fibers will cause only a weak response in the muscle, but simultaneous stimulation of all 100 fibers will cause a strong contraction. Obviously, any number of fibers between 1 and 100 can be stimulated at a time, giving any one of 100 different strengths of muscle contraction.

Temporal summation means changing the strength of a signal by sending a large or small number of impulses along the same fiber per second. If one impulse is transmitted each second, only a weak effect usually results, but if 5, 15, 25, 75 or more impulses are transmitted per second, the strength of the effect becomes progressively greater.

Ordinarily the nerve trunk transmits signals of different strengths by a combination of the spatial and temporal methods. That is, when a strong signal is to be transmitted, large numbers of fibers are utilized and large numbers of impulses are transmitted along each fiber. When a weak signal is to be transmitted, fewer fibers are used and fewer impulses are transmitted.

TRANSMISSION OF IMPULSES BY TISSUES OTHER THAN NERVE FIBERS

Transmission by skeletal muscle fibers. Skeletal muscle fibers transmit impulses exactly as nerve fibers. The normal velocity of transmission in skeletal muscle fibers is about 5 meters per second in contrast to 50 to 100 meters per second in the very large nerve fibers and 0.5 meter per second in the very small fibers. Because the nerve fibers supplying skeletal muscles are a large type, carrying impulses at about 50 meters per second, a signal travels from the brain to the muscle extremely rapidly but then decreases in velocity about 10-fold as it goes into the muscle itself.

Transmission in heart muscle and smooth muscle. Transmission of impulses in heart muscle and smooth muscle occurs like that in skeletal

muscle, but at still lower velocities—about 0.3 meter per second in the heart and only a centimeter or so per second in smooth muscle. In heart muscle and in many smooth muscle masses the fibers interconnect with each other to form latticeworks so that stimulation of any one fiber always causes the impulse to travel over the entire muscle mass, resulting in complete contraction of the whole muscle rather than only part of it. Another difference between these two types of muscle and skeletal muscle is the duration of the refractory period, which in heart muscle is about 0.3 second and in smooth muscle sometimes several seconds. A second impulse cannot stimulate the muscle again for this long period of time. Also, as long as the membranes remain depolarized, the muscle fibers remain contracted. Therefore, the duration of contraction of both heart and smooth muscle is unusually long in comparison with that of skeletal muscle.

Transmission of impulses over the membranes of glandular and other cells. Because most glandular secretions of the body are elicited by nerve impulses, it is believed by some physiologists that nerve impulses on reaching the glands cause secondary impulses to travel over the membranes of the secretory cells, this eliciting the secretion. It is also possible that the white blood cells and other cells that exhibit ameboid motion are controlled by electrical potentials that spread over their bodies. Perhaps even the phenomenon of chemotaxis results from membrane action potentials initiated by the chemotactic substance. And the migration of fibroblasts into areas that need tissue repair possibly is also controlled by electrical potentials spreading over their surfaces.

THE NEUROMYAL JUNCTION

The *neuromyal junction* is the connection between the end of a large, myelinated nerve fiber and a skeletal muscle fiber. In general, each skeletal muscle is supplied with at least one neuromyal junction and occasionally with several. Figure 198 illustrates a neuromyal junction, showing a nerve fiber passing beneath the muscle membrane, called the *sarcolemma,* and then spreading to form many branches, called *hypolemmal axons.* These axons in turn end in club-like feet called *sole feet.* The entire nerve ending, including the hypolemmal axons and the sole feet, is called the *endplate.* Between the sole feet and the *plasma membrane* of the muscle filter is a small space. It is the plasma membrane that transmits the impulse over the muscle fiber.

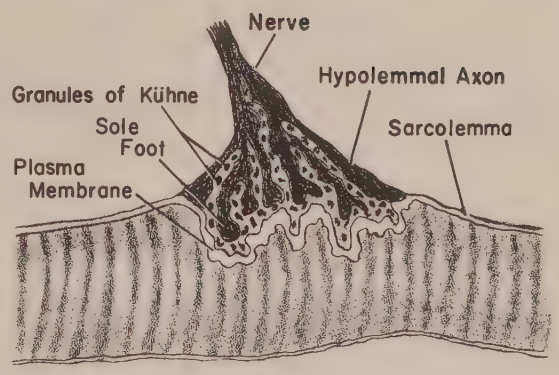


Figure 198. The neuromyal junction.

Transmission of the Impulse at the Neuromyal Junction

Secretion of acetylcholine. When a nerve impulse reaches the neuromyal junction, the endplate secretes a chemical called *acetylcholine* into the space between the sole feet and the plasma membrane. The acetylcholine then acts on the plasma membrane to increase its permeability. The increased permeability in turn allows instantaneous leakage of sodium to the inside of the fiber, which alters the membrane potential beneath the endplate. This change in potential is called the *endplate potential.* If the endplate potential becomes great enough, it will cause an action potential to travel in both directions along the muscle fiber. In turn, the action potential traveling along the fiber causes muscle contraction, as will be explained in Chapter 21.

Destruction of acetylcholine by cholinesterase. If the acetylcholine secreted by the endplate should remain in contact with the membrane indefinitely, the muscle fiber would transmit a continuous succession of impulses. However, in all the fluids of the body, and especially in the neuromyal junction itself, is a substance called *cholinesterase,* which enzymatically splits acetylcholine into acetic acid and choline in about $\frac{1}{500}$ second. Therefore, almost immediately after the acetylcholine has stimulated the muscle fiber the acetylcholine itself is destroyed. This allows the membrane to repolarize and to become ready for stimulation again as soon as a new nerve impulse approaches.

The acetylcholine mechanism at the neuromyal junction provides an amplifying system that allows a very weak nerve impulse to stimulate a very large muscle fiber, for the amount of electrical current generated by the nerve fiber is not enough by itself to elicit an impulse in the muscle

fiber. Instead, the secreted acetylcholine causes the muscle fiber to generate its own impulse. Thus, each nerve impulse actually comes to a halt at the neuromyal junction, and in its stead an entirely new impulse begins in the muscle.

Effect of Drugs on the Neuromyal Junction

Effect of curare. *Curare* blocks transmission of the impulse at the neuromyal junction by making the muscle membrane resistant to the action of acetylcholine. Therefore, the acetylcholine is unable to stimulate the muscle, and the muscle becomes paralyzed.

Neostigmine and neostigmine-like drugs. *Neostigmine* and similar drugs such as *physostigmine* have an opposite effect to that of curare, for they enhance the transmission of impulses at the neuromyal junction. They do so by inhibiting the action of cholinesterase. As a result, the acetylcholine secreted by the endplate is not destroyed but instead accumulates in progressively larger and larger quantities. Eventually, so much acetylcholine is present that every time the muscle fiber repolarizes it is immediately depolarized again. As a result, the muscle fiber receives a succession of impulses and becomes spastic.

Another drug that has similar effects to those of neostigmine and physostigmine is diisopropyl fluorophosphate, but this drug has a much more prolonged effect. Instead of simply inhibiting the cholinesterase for a few hours, it destroys the cholinesterase, and several weeks are required for resynthesis of new cholinesterase. Therefore, diisopropyl fluorophosphate can cause spastic paralysis for several weeks. This substance is one of the dreaded "nerve"gases that possibly will be used in some future war.

Myasthenia Gravis

Sometimes a person has very poor transmission of impulses at the neuromyal junction; this condition is called *myasthenia gravis*. Though the exact cause of this condition is unknown, it could result from any one of three abnormalities: (1) deficient acetylcholine secretion, (2) too rapid destruction of acetylcholine by cholinesterase, or (3) depressed responsiveness of the muscle fiber membranes to acetylcholine. Even though the cause is not known, treatment with neostigmine or with some other drug that prevents the destruction of acetylcholine is often very beneficial. These allow acetylcholine to accumulate in the neuromyal junction and to react with the muscle fiber membrane for a prolonged time. As a result, persons almost totally paralyzed by myasthenia gravis can sometimes be returned to complete normality within less than one minute after a single intravenous injection of neostigmine.

REFERENCES

Fernandez-Moran, H., and Brown, R.: *Submicroscopic Organization and Function of Nerve Cells.* New York, Academic Press, 1958.

Frank, K., and Fuortes, M. G. F.: Excitation and conduction. *Ann. Rev. Physiol.* 23:357, 1961.

Hodgkin, A. L., and Horowicz, P.: Movements of Na and K in single muscle fibers. *J. Physiol.* 145:405, 1959.

Hodgkin, A. L., and Huxley, A. F.: Resting and action potentials in single nerve fibres. *J. Physiol.* 104:176, 1945.

Rosenblueth, A.: *The Transmission of Nerve Impulses at Neuroeffector Junctions and Peripheral Synapses.* New York, John Wiley & Sons, Inc., 1950.

Shanes, A. M.: Electrochemical aspects of physiological and pharmacological action in excitable cells. *Pharmacol. Rev.* 10:59, 1958; 10:165, 1958.

Stampfli, R.: Conduction and transmission in the nervous system. *Ann. Rev. Physiol.* 25:493, 1963.

Terzuolo, C. A., and Edwards, C.: Excitation and synaptic transmission. *Ann. Rev. Physiol.* 24:325, 1962.

MUSCLE PHYSIOLOGY

21

All physical functions of the body involve muscle activity. These functions include skeletal movements, contraction of the heart, contraction of the blood vessels, peristalsis in the gut, and many more. Three different types of muscle are responsible for these activities: skeletal muscle, cardiac muscle, and smooth muscle, all of which have some characteristics in common. For instance, the contractile process is the same in each, but, on the other hand, their reactivity, their durations of contraction, and other features differ greatly and are especially adapted in each type of muscle for the job to be performed.

SKELETAL MUSCLE

The skeletal muscles cause movements of the skeleton and, therefore, are responsible for movement of the different parts of the body. These muscles are controlled by nerve impulses transmitted from the brain, and by means of muscular contractions a person walks, performs skilled movements with his hands, positions his body, talks, breathes, makes facial expressions, and performs any other physical activity.

Physiologic Anatomy of Skeletal Muscle

Figure 199 illustrates a single muscle, showing schematically that the *muscle belly* is made up of thousands of individual *muscle fibers* attached at each end to *muscle tendons.* When stimulated,

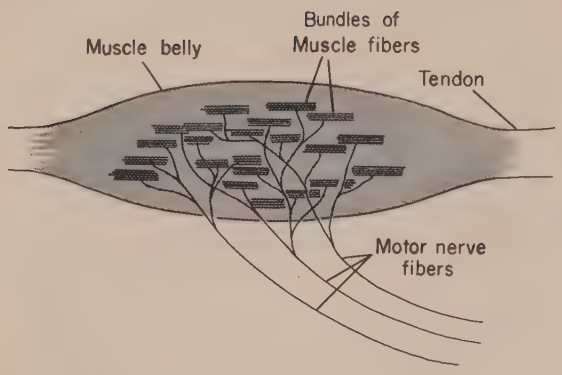


Figure 199. A muscle, showing the distribution of nerve fibers supplying three motor units.

the muscle fibers contract, and the force of contraction is transmitted to the bones through the tendons.

Figure 200 illustrates cross-sectional and longitudinal views of a single muscle fiber. Each fiber is between 10 and 100 microns in diameter, and it varies from a few millimeters to 50 centimeters in length. The longitudinal view shows dark and light bands along the fiber, which is characteristic of skeletal and cardiac muscle but not of smooth muscle. A pair of these bands is called a *sarcomere.* Each muscle fiber contains a thousand or more longitudinal *myofibrils* arranged in bundles within the fiber, as shown in the cross-sectional view in Figure 200. The myofibrils in turn are composed of many small fibrils composed of *actin-tropomyosin* and *myosin* protein molecules which are actual contractile elements of the muscle.

Figure 201 illustrates several sarcomeres, showing by the upper diagram the dark and light bands and by the lower diagram the actual arrangement of actin and myosin filaments in the fiber. The dark band is called the *A band* or *anisotropic band* and the light band is called the *I band* or *isotropic band.* Note that it is the arrangement of the actin and myosin filaments in the fiber that makes the light and dark bands, the anisotropic band corresponding to the myosin filaments and the isotropic band corresponding to the segment of the fiber between successive myosin filaments. Note also that the myosin and actin filaments interdigitate. During muscle contraction, the actin filaments slide inward among the myosin filaments, thus resulting in shortening of the muscle as is explained below.

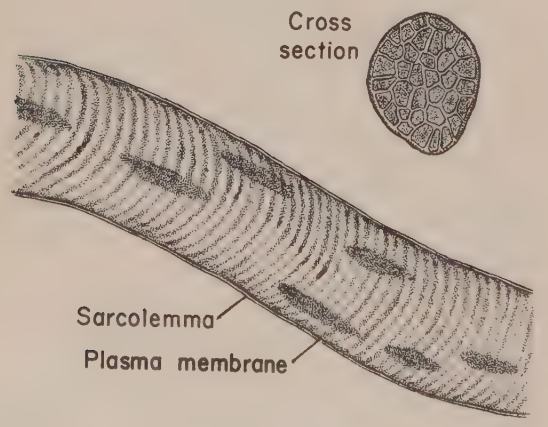


Figure 200. Longitudinal and cross-sectional views of a skeletal muscle fiber.

The muscle fiber is enclosed by a membrane called the *sarcolemma,* and immediately beneath the sarcolemma is a *plasma membrane* which is the true cell membrane of the muscle fiber and which is capable of transmitting action potentials in the same manner that the nerve membrane transmits action potentials.

Skeletal Muscle Contraction

Excitation of the Skeletal Muscle Fiber

When a nerve impulse reaches the neuromyal junction, acetylcholine is released as explained in the previous chapter, and this acts on the plasma membrane to make it permeable to sodium ions. The rapid influx of sodium ions creates an electrical potential at the neuromyal junction called the *endplate potential* which, if strong enough, initiates an impulse traveling in both directions along the membrane of the muscle fiber. The impulse travels at a velocity of approximately 5 meters per second. Therefore, after stimulation of a 10 cm. muscle fiber by a neuromyal junction in its middle, the impulse will reach both ends of the fiber in approximately $\frac{1}{100}$ second. It can be seen, therefore, that an impulse excites the entire fiber in a very short period of time.

Function of calcium in initiating muscle contraction. Passage of the action potential along the muscle fiber membrane is accompanied by an increase in the permeability of the membrane. In the previous chapter it was pointed out that under these conditions sodium ions flow profusely to the inside of the fiber. In addition to sodium, however, a reasonable amount of calcium also enters the fiber, and recent experiments have shown that it is these calcium ions entering the fiber that cause the actual contraction, as is explained below.

The Contractile Process

Figure 202 illustrates the mechanism of skeletal muscle contraction, showing in the upper drawing the relaxed state of the fiber and in the lower drawing the contracted state. This figure illustrates that neither the myosin nor the actin filaments decrease in length but instead that the actin filaments simply slide like pistons inward among the myosin filaments. The precise manner in which this sliding process is effected at the onset of muscular contraction is still somewhat conjectural, but many of the steps have been discovered. The process is believed to be the following:

When calcium ions suddenly appear in the fluids of the muscle fiber, some of these combine loosely with myosin molecules to give them the property of an enzyme called *ATPase,* meaning by this that the *activated myosin* can now react with adenosine triphosphate (ATP) to remove its energy. Attached to the actin filaments is a large amount of ATP, and extending from the myosin filaments toward the actin filaments are numerous *cross-bridges,* which can be seen in Figure 202. The activated myosin ATPase releases the energy from the ATP on the actin filaments, probably by interaction between the cross-bridges and the actin, which in turn is utilized to pull the actin filaments inward among the myosin filaments. It is presumed that the chemical reactions involved in this process create electrostatic charges between the myosin and actin filaments which cause them to attract each other and thereby promote the sliding process.

Relaxation of the muscle fiber—"the relaxing factor." Once calcium ions have been released to the interior of the muscle fiber, energy will continue to be liberated from adenosine triphosphate, and the contracted state of the muscle will continue to be maintained until either the muscle wears out from complete breakdown of all its available adenosine triphosphate or until the calcium ions are removed from the fluids. Fortunately, present everywhere in the cytoplasm surrounding the filaments is a substance called *relaxing factor.* This substance has a natural affinity for calcium ions, which combine with it within a few hundredths of a second. Therefore, in effect, an action potential causes a "pulsation" of calcium ions inside the fiber, their concentration building up very rapidly during the action potential and then falling off to zero within a few hundredths of a second. It is during this pulsation that the muscle contracts; at the end of the pulsation, relaxation takes place.

If prolonged contraction is to take place, calcium ions must continually move into the muscle fiber. This is accomplished by repetitive passage of a series of action potentials one after another over the membrane of the fiber. With each action potential a new pulsation of calcium ions is delivered to the interior of the fiber, and if these pulsations follow one upon the other very rapidly, a constant quantity of calcium ions can be maintained inside the fiber. Furthermore, the more rapid the repetition of the action potential, the greater becomes the concentration of calcium ions, and in turn the greater becomes the force of muscle contraction.

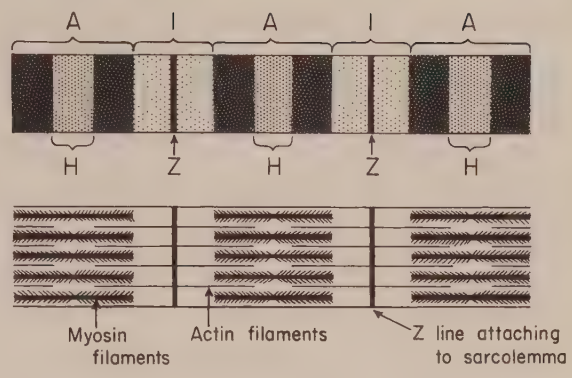


Figure 201. *Above:* Schematic diagram of the light and dark areas in a myofibril. *Below:* Arrangement of the myosin and actin filaments in the sarcomeres.

Once calcium ions have combined with the relaxing factor, they must eventually be removed to the exterior of the cell, but, unfortunately, the method by which this occurs is yet unknown.

The energy for contraction. The immediate source of energy for muscle contraction is the adenosine triphosphate attached to the actin

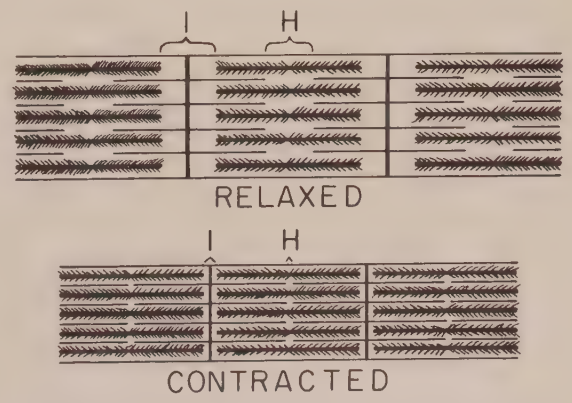


Figure 202. The relaxed and contracted states of a myofibril, showing sliding of the actin filaments into the channels between the myosin filaments.

filaments. Adenosine triphosphate (ATP) is a so-called "high-energy compound" that functions in all cells of the body to provide a ready source of energy for almost any type of cellular function. When one mole of adenosine triphosphate splits to form a mole of phosphoric acid and a mole of adenosine diphosphate (ADP), about 7000 calories of energy are released. It is this energy that in some yet unknown way creates the forces between the myosin and actin filaments that cause muscle contraction.

Once ATP has been split into ADP and phosphoric acid, these substances must be reconverted into ATP for reuse at a later time. This is accomplished by the metabolism of foodstuffs. For instance, the metabolism of glucose with oxygen to form carbon dioxide and water releases about 680,000 calories of energy per mole of glucose. This energy is used to synthesize 38 moles of ATP from ADP. The ATP can also be replenished by energy from fat or protein metabolism as well. Thus, increased metabolism in a muscle fiber after contraction is over replenishes the chemical stores that are used up during the contractile process. Also, energy must be used to expulse the calcium and sodium ions that enter the fiber during contraction.

HEAT LIBERATED BY THE CONTRACTILE PROCESS. Unfortunately, not all of the energy stored in the form of adenosine triphosphate is converted into muscle work, and, likewise, not all of the energy stored in the glucose molecule is utilized in the formation of ATP. Instead, because of a certain degree of *inefficiency* of all energy transfer processes in the body, much energy is lost in the form of heat. Figure 203 illustrates that heat is lost from a muscle both during and after the period of contraction. This heat is divided into four different parts: (1) heat caused by the actual shortening of the muscle, called the *heat of shortening,* (2) heat caused by the chemical processes required to initiate and to maintain contraction, called the *heat of activation and maintenance,* (3) heat caused by viscous changes and other physical changes in the muscle during relaxation, called the *heat of relaxation,* and (4) heat required to replenish the ATP and to remove calcium and sodium from the interior of the fiber, called the *heat of recovery.*

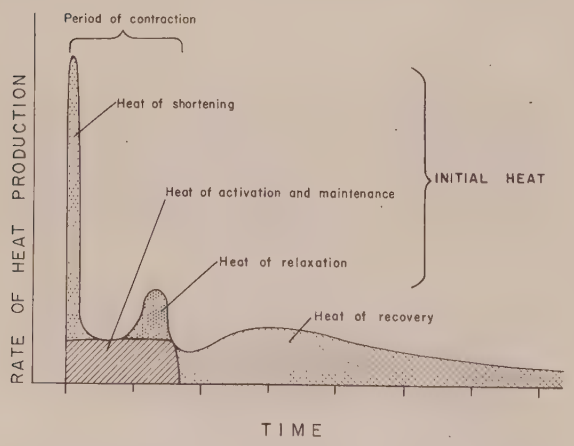


Figure 203. Heat production during and following muscle contraction, showing the proportions of heat resulting from different phases of the contractile process.

EFFICIENCY OF MUSCLE CONTRACTION. The *efficiency* of an energy exchange process is the percentage of the input energy that becomes output energy. In muscle contraction, the input energy is the potential energy of the nutrients used by the muscle fibers for contraction, and the output energy is the work output of the muscle. Because of inefficiencies of energy transfer in the muscle fibers, the percentage of input energy that eventually becomes work output, even under optimal conditions, is only 20 to 25 per cent; the other 75 to 80 per cent becomes heat. Under some conditions, such as when standing absolutely still, a person has no work output from his muscles, and, yet, a large amount of nutrients is being utilized to maintain the tension in his muscles. Under these conditions the efficiency of muscle contraction is absolutely zero.

The all or none principle. The all or none principle applies to contraction of the skeletal muscle fiber in exactly the same way that it applies to transmission of an impulse over a nerve fiber. Once a muscle fiber has become excited and an action potential has begun to spread along its membrane, the entire fiber responds. That is, a muscle fiber cannot be "partially" excited; either it does not become excited at all or it becomes excited in its entirety.

The all or none principle, however, does not imply that the strength of contraction is the same each time the fiber is stimulated. The contractile elements of the muscle fiber are sometimes weaker than at other times because of fatigue, lack of nutrients, or other causes, so that even though the entire fiber responds, the strength may vary from time to time. In essence, then, the all or none principle means that when a muscle fiber is excited the entire fiber contracts to the full extent of its immediate ability.

The Motor Unit

Figure 199 showed three separate nerve fibers entering a muscle. Actually this is merely a figurative representation, because several hundred to

several thousand nerve fibers enter each muscle. The figure also showed that each of the nerve fibers divides and spreads throughout the muscle belly to terminate on many different muscle fibers. On the average, a single motor nerve fiber innervates about 150 muscle fibers, which means that stimulation of one nerve fiber will cause contraction of 150 muscle fibers at the same time. All the fibers innervated by the same nerve fiber are called a *motor unit* because they always contract in unison.

It is especially important that all of the muscle fibers in a motor unit do not necessarily lie side by side, but instead are divided into many bundles of only a few fibers each, spread throughout the muscle belly. Because of this, stimulation of the motor unit causes a weak contraction in a broad area of the muscle rather than a strong contraction at one specific point.

Those muscles that must control very fine movements usually have only a few muscle fibers in each motor unit, which means that the ratio of nerve fibers to muscle fibers is very high. For instance, the ocular muscles, which must control the extremely discrete movements of the eyes, have no more than 10 muscle fibers to each motor unit. On the other hand, postural muscles, which usually exhibit only very gross movements, may have as many as 200 or more fibers to the motor unit.

The Muscle Twitch

One of the laboratory methods for studying muscle contraction is to elicit a *muscle twitch*. To do this, a single instantaneous stimulus is applied to the nerve supplying the muscle. The duration of the resulting contraction is between ⅒ and ⅓₀₀ second, depending on the type of muscle, which is so short that is appears to be a "twitch."

Isometric and isotonic contraction. Figure 204 shows the contraction of a muscle under two different conditions. To the left the muscle lifts weights in a pan, becoming shorter in the process. The total amount of weight applied to the muscle is always the same, for which reason the contraction is called *isotonic*, which means "same force." To the right the muscle is attached between a solid bar and a tight torsion spring, which cannot move more than about ⅟₁₀₀₀ inch. That is, the muscle belly is suspended between two rigid points. Stimulation of the muscle under these conditions causes it to tighten but not to shorten significantly, and the contraction is called *isometric*, meaning "same length."

The characteristics of isometric and isotonic

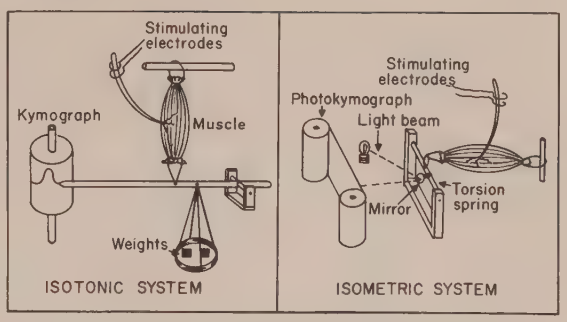


Figure 204. Methods for recording isotonic and isometric muscle twitches.

muscle contraction are somewhat different. The reasons for this are, first, the isometric system has no inertia while the isotonic system does, and, second, during isotonic contraction the shape of the muscle must change so that it can shorten, whereas during isometric contraction the muscle does not need to change shape but only to create force. Therefore, without inertia and without the necessity for changing shape, the isometric muscle twitch usually has a much shorter duration than the isotonic twitch. The duration of the isotonic twitch depends to a great extent on the inertia of the object to be moved and on the degree of muscle shortening that occurs. In expressing relative abilities of different muscles to contract, the isometric muscle twitch is the usual criterion employed, because its characteristics are dependent only on intrinsic characteristics of the muscle and not at all on extrinsic factors.

In the human body, muscle contraction is of both the isometric and isotonic types. When one is simply standing he tenses his leg muscles to maintain a fixed position of the joints. This is isometric contraction. On the other hand, when he is walking and moving his legs, or when he is lifting his arms, the contraction is more of the isotonic type.

Duration of contraction of different skeletal muscles. Figure 205 shows recordings of isometric contractions by different muscles. The dashed curve of the figure shows the duration of depolarization of the fiber membrane caused by the action potential traveling over the muscle fiber. This is the period when calcium ions are shifting into the fiber. Immediately thereafter the contraction begins. The isometric contraction of an ocular muscle lasts for about ⅟₁₀₀ second while that of a gastrocnemius muscle lasts about ³⁄₁₀₀ second. It is evident, then, that different skeletal muscles have different durations of contraction. The ocular muscles, which must cause extremely rapid movement of the eyes from one

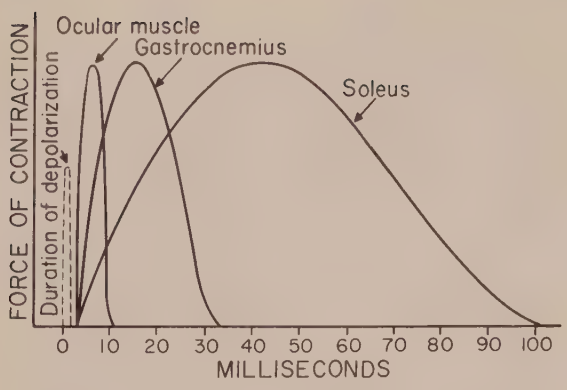


Figure 205. Isometric muscle twitches of ocular, gastrocnemius, and soleus muscles, illustrating the different durations of contraction.



Figure 207. The lever system activated by the biceps muscle.

position to another, contract more rapidly than almost any other muscle. The gastrocnemius muscle must contract moderately rapidly because it is used in jumping and in performing other rapid downward movements of the foot. The soleus need not contract rapidly at all, because it is used principally to perform the prolonged movements for support of the body against gravity.

Effect of Initial Muscle Length on the Force of Contraction

The length to which a muscle is stretched before it contracts makes considerable difference in its force of contraction. When the length is much less than normal its force of contraction is greatly weakened, and, also, when it has been stretched far beyond its normal limits, it fails to contract with as much force as would otherwise be possible. Figure 206 illustrates these effects, showing that a muscle in its normal stretched state will usually contract with the greatest possible force.

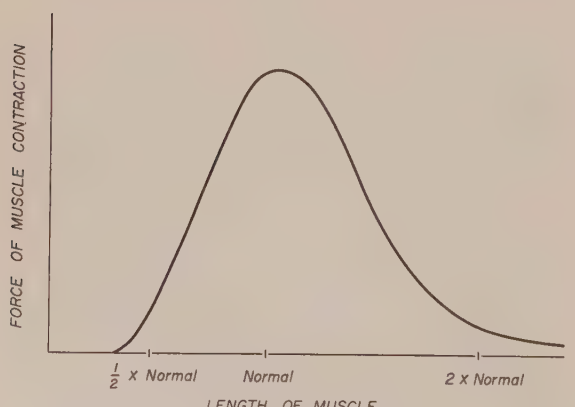


Figure 206. Effect of the initial length of a muscle on the contractile force developed following muscle excitation.

Fortunately, the normal length of a muscle in its most elongated position is almost exactly optimal for maximal strength of contraction. For instance, when the biceps is at its normal full length, it contracts with its greatest force, whereas, as it progressively shortens, its strength of contraction decreases.

The lever systems of the body. Other factors that determine the force of a movement are (a) the manner in which the contracting muscles are attached to the skeletal system and (b) the structure of the joint at which movement will occur. Figure 207, as an example, illustrates movement of the forearm by biceps contraction. The fulcrum of the lever system is at the elbow, and the attachment of the biceps is approximately two inches in front of the fulcrum. If we assume that the total length of the forearm lever is about 14 inches, one immediately sees that the force of contraction of the biceps must be at least 7 times as great as the force of movement of the hand. Thus, if the hand is to lift an object that weighs 50 pounds, the total force of contraction of the biceps would have to be about 350 pounds. One can readily see from this tremendous force why muscles sometimes pull their tendons out of the bone substance.

Every muscle of the body has its own peculiar shape and length that suits it to its particular function. For instance, the muscles of the buttocks are extremely broad but do not contract a long distance. They provide tremendous force for movement at the hip joint, and even a very slight distance of movement at this joint can cause tremendous movement of the foot. At the other extreme, some of the muscles of the anterior thigh are very long and can shorten almost a foot, pulling the lower leg upward at the knee joint and

flexing the upper leg at the hip joint at the same time.

The study of different types of muscle lever systems and their movements is called *kinesiology;* this is a very important phase of human physioanatomy.

Different Strengths of Muscle Contraction—Summation

In performing the different functions of the body, it is quite important that each muscle be able to contract with varying degrees of strength. This is accomplished by "summing" the contractions of varying numbers of muscle fibers at once. When a weak contraction is desired, only a few muscle fibers are contracted simultaneously. When a strong contraction is desired, a great number of fibers are contracted at the same time. In general, the different "gradations" of muscle contraction are achieved by two different methods of summation called *multiple motor unit summation* and *wave summation.*

Multiple motor unit summation. If only one motor unit contracts at a time, the muscle contraction will be weak, but if all the motor units of the muscle contract simultaneously, the strength of contraction will be maximal. All gradations of muscle contraction between minimal and maximal can be attained by varying the number of motor units contracting. This effect is illustrated in Figure 208, which shows the force of various numbers of motor units contracting simultaneously.

Wave summation. Wave summation occurs when each muscle fiber contracts many times in rapid succession, the contractions occurring closely enough together so that a new contraction occurs before the previous one is over. Each succeeding contraction adds to the force of the preceding one, increasing the overall strength of contraction. This effect is illustrated in Figure 209. To the left the contractions occur at a rate of five per second, which is long enough apart so that they do not summate. But, by the time the rate has reached 10 to 15 times per second, successive contractions occur before previous ones are over. As a result, the contractions add to each other, causing an increased force of contraction. In this figure the force of contraction increases progressively until the rate of stimulation reaches about 55 times per second. Some of the rapidly acting muscles, such as the ocular muscles, require as many as 400 to 600 stimulations per second for maximum contraction to take place.

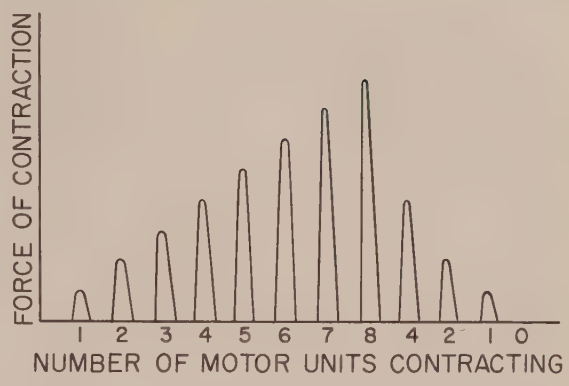


Figure 208. Multiple motor unit summation, illustrating the progressive increase in contractile force caused by increasing numbers of motor units contracting simultaneously.

TETANIZATION. Observing Figure 209 once again, it can be seen that individual contractions are no longer evident after the rate of stimulation has risen above about 35 per second, but instead a smooth, continuous contraction occurs. In other words, the muscle twitches become fused into a single prolonged contraction. This effect is called *tetanization.* The frequencies at which different muscles will tetanize depend on the durations of single contractions. In the soleus muscle tetanization occurs at a frequency of about 20 contractions per second because the soleus is a slowly contracting muscle. Likewise, the gastrocnemius tetanizes at about 50 stimuli per second because it is a moderately rapidly contracting muscle.

Asynchronous summation. In normal muscle contraction the different gradations of contraction are attained by utilizing a mixture of multiple motor unit summation and wave summation. This is effected by contraction of all the different motor units of a muscle belly one at a time, but rapidly

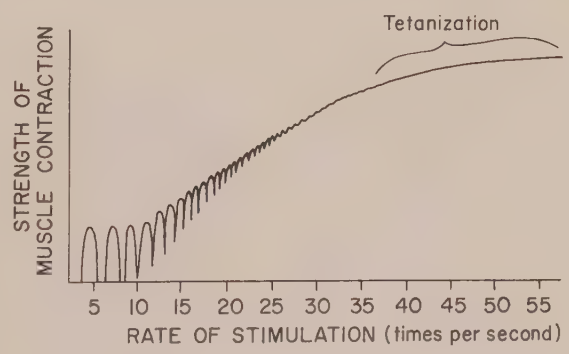


Figure 209. Wave summation, showing progressive summation of successive contractions as the rate of stimulation is increased. Tetanization occurs when the rate of stimulation reaches approximately 35 per second, and maximum force of contraction occurs at approximately 50 per second.

enough in succession so that the muscle tension is always of a tetanic type rather than of a twitching type. For a weak contraction each motor unit contracts only two to three times per second, but the contractions are spread out one after another among the different motor units, rather than all units contracting at the same time. In this way some motor units are always contracting, assuring constant though weak tension on the muscle. When a strong contraction is desired each motor unit contracts 50 or more times per second.

Asynchronous summation is caused by neuronal circuits in the spinal cord that automatically distribute the impulses evenly and sequentially among the different nerve fibers to a muscle.

Tone of skeletal muscle. The spinal cord continually transmits a few impulses to all skeletal muscles most of the time, which maintains slight tension in the muscles and keeps them from becoming flabby. This is called *muscle tone,* and it keeps the muscle "taut" so that it will react to a strong stimulus far more rapidly than it could otherwise.

The tone of skeletal muscle may increase or decrease depending on the physiologic activity of the body. When a person is in a state of anxiety, excitement, fear, or certain other emotional conditions, the number of tone impulses from the central nervous system becomes intensified, causing the muscles to become increasingly taut so that they respond to other stimuli with extreme rapidity. This is the basis of the jumpiness of the so-called "nervous" individual. On the other hand, during sleep the number of tone impulses becomes almost nil, allowing the muscles to relax completely.

Effect of Activity on Muscular Development

Exercise and hypertrophy. The more a muscle is used, the greater becomes its size and strength. This is a very valuable characteristic of skeletal muscle, though its cause is unknown. If a person continually uses certain muscles for performing special duties, these muscles become larger and stronger than his other muscles; or, if he uses all his muscles to an excessive amount, all of them become greatly strengthened. The physical enlargement of the muscles is called *hypertrophy;* examples of this are (1) the intense muscular development of weight lifters, (2) the great hypertrophy of the leg muscles in skaters, (3) the enlargement of the arm and hand muscles of carpenters, and (4) the enlargement of the thigh muscles of runners.

Associated with muscle hypertrophy is usually an increase in the efficiency of muscular contraction, for the hypertrophied muscle stores increased quantities of glycogen, fatty substances, and other nutrients, and the number of contractile fibrils increases. All these cause the efficiency of the contractile process to increase so that the percentage of energy lost as heat becomes considerably less in the athlete than in the non-athlete.

It should be noted that weak muscular activity, even when sustained over long periods of time, does not result in significant muscle hypertrophy. Instead, hypertrophy is mainly the result of forceful muscle activity even though the activity might occur only a few minutes each day. For this reason, strength can be developed in muscles more rapidly by using "resistive exercises" than by prolonged mild exercise.

Muscle denervation and atrophy. When the nerve supply to a muscle is destroyed the muscle begins to *atrophy*—that is, the muscle fibers begin to degenerate. In about six months to two years the muscle will have atrophied to about one-fourth normal size, and the muscle fibers will have been replaced mainly by fibrous tissue.

For some reason nerve stimulation of a muscle keeps the muscle tissue alive. Even when a person does not use his muscles to a great extent, the weak tone impulses are still sufficient to maintain a relatively normal muscle, but without these impulses the muscle fibers soon atrophy entirely. Perhaps this effect is caused by nutritional changes in the denervated muscle, for action potentials traveling down the fiber membrane alter its permeability markedly, which might be necessary for appropriate transfer of nutrients through the membrane. The integrity of denervated muscle fibers can be maintained quite satisfactorily by daily electrical stimulation. The action potentials produced in this manner take the place of the nerve induced potentials, and the muscle fibers do not atrophy.

CARDIAC MUSCLE

Referring back to Figure 98 of Chapter 12, one can see that cardiac muscle fibers have the same striated appearance as that seen in skeletal muscle. The only significant difference between the appearance of the two types of muscle is that cardiac fibers are arranged in a syncytium—that is, they interconnect with each other. The heart is composed of two separate syncytiums. One of these forms the walls and septum of the atria, and the

other forms the walls and septum of the ventricles. When a single fiber of either of the syncytiums is stimulated, all the connecting muscle fibers in this syncytium become stimulated also. Therefore, in applying the all or none principle to the contraction of cardiac muscle, one does not speak of each muscle fiber separately but instead of the syncytiums. That is, a stimulus applied to the heart either causes the muscle syncytium to contract entirely or not at all. Impulses are transmitted from the atria into the ventricles through the Purkinje system, which was described in Chapter 12. Because of this connecting link, an impulse beginning anywhere in the atrial syncytium spreads on into the ventricles, causing the entire heart to contract.

Rhythmicity of cardiac muscle. One of the major physiologic differences between skeletal muscle and cardiac muscle is that skeletal muscle normally contracts only when stimulated by a nerve, whereas cardiac muscle contracts rhythmically approximately 72 times per minute, never stopping. The contraction is caused by an inherent property of all cardiac muscle to discharge its membrane potential every time the potential builds up. That is, every time the membrane repolarizes after a previous cardiac impulse, the membrane, for yet inexplicable reasons, suddenly becomes permeable again, allowing transmission of a new action potential over the entire heart. This event occurs over and over again approximately once every $^{8}/_{10}$ second.

When skeletal muscle is denervated, it also exhibits rhythmicity beginning approximately two to three days after denervation and continuing until the muscle becomes atrophic. The rate of contraction is ten or more times as rapid as that of cardiac muscle, and the resulting movements are called *fibrillations*. This fact simply illustrates that the phenomenon of rhythmicity is probably a fundamental property of all muscle. In skeletal muscle it has been subordinated to nervous control, while in the heart the rhythmic contractions are of primary importance and nervous control is of secondary importance.

Prolonged depolarization and prolonged contraction of cardiac muscle. Another difference between cardiac muscle and skeletal muscle is the duration of depolarization and contraction. After stimulation, the membrane of cardiac muscle remains depolarized for about $^{3}/_{10}$ second in contrast to about $^{1}/_{500}$ second in skeletal muscle, and as long as the membrane remains depolarized the contractile process of the actin and myosin molecules is maintained. Therefore, the period of a single cardiac muscle contraction is about $^{3}/_{10}$ second. This prolonged contraction is a very important characteristic of cardiac muscle, because, for the heart to operate properly, its contraction must last long enough to force blood from its chambers.

SMOOTH MUSCLE

Most of the internal organs of the body contain *smooth muscle*, the third type of muscle. The name is derived from the fact that this muscle does not have striations similar to those in skeletal and cardiac muscle.

The smooth muscle of each internal organ is at least a little different from that in any other organ. In most organs the smooth muscle fibers are so tightly bound to each other that, from a functional point of view, they form syncytiums similar to that of cardiac muscle. Syncytial smooth muscle is illustrated to the left in Figure 210. On the other hand, some smooth muscle fibers are discrete entities similar to the fibers of skeletal muscle. These are illustrated to the right in Figure 210.

Smooth muscle fibers may be only 20 to 40 microns in length, or they may be several millimeters long, depending on the function to be performed. Quite frequently they are arranged in large sheets rolled into tubes or spheres to form such organs as the gut, the bladder, blood vessels, and many others. Other smooth muscle fibers are arranged like a ring, as around the pupil of the eye, forming a sphincter that can decrease the size of an opening. They occasionally, also, are arranged in tufts similar to the muscle bellies of skeletal muscle. These are seen especially at the bases of hairs where several smooth muscle fibers

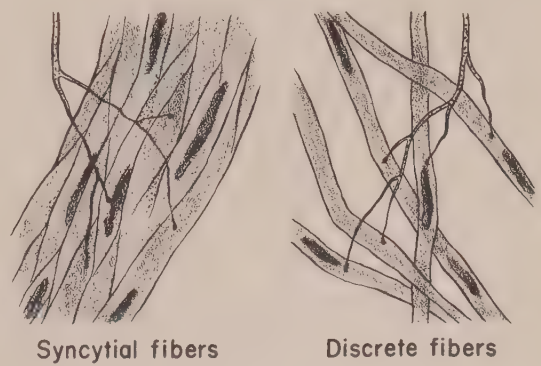


Figure 210. Syncytial and discrete organization of smooth muscle fibers.

form each pilo-erector muscle that causes the hairs to stand on end when a person becomes cold or frightened. In summary, except for the movements of the skeleton and the pumping action of the heart, all the physical activities of the body are performed by smooth muscle. This type of muscle is found in almost all internal organs, and its specific characteristics vary with the function to be performed.

The contractile process in smooth muscle. Smooth muscle contains myosin and actin filaments the same as skeletal and cardiac muscle, but the sarcomeric arrangement that causes the striated appearance in these other types of muscle is absent. Smooth muscle also contains the same ionic constituents as the other two types of muscles. Therefore, it is believed that contraction in smooth muscle occurs in the same way as in both skeletal and heart muscle, by interaction between myosin and actin filaments initiated by calcium ions. Yet because there is no piston-like arrangement between the two types of filaments in smooth muscle such as one finds in skeletal muscle, some other *mechanical* explanation for the contraction besides the piston movement must be found. Though an exact explanation is unknown, it is assumed that when myosin becomes activated, the actin filaments, however they happen to lie at the moment, are made to slide in among the myosin filaments. Since these filaments are oriented linearly along the muscle fiber, any such sliding process must of necessity shorten the muscle. It is likely that this problem will be solved soon, now that the electron microscope can reveal structural details of the myofibrils.

Rhythmicity of smooth muscle. Most smooth muscle exhibits rhythmicity similar to that in cardiac muscle. This is especially true of the syncytial types, which comprise perhaps 99 per cent of all smooth muscle. The precise cause of the rhythmicity is not known, but it is presumed that the membrane potential finally reaches a voltage that sets off an impulse itself. Once the impulse is started in any single fiber of the syncytium, it travels throughout the entire muscle mass.

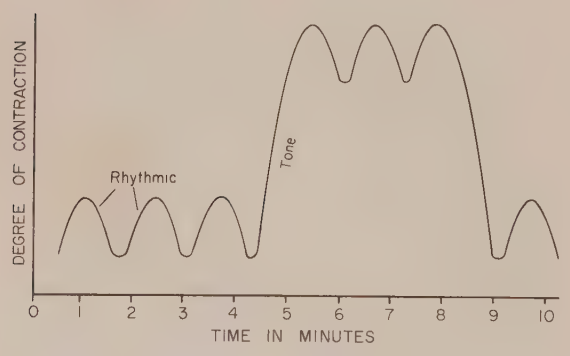


Figure 211. Rhythmic and tonic contractions of smooth muscle.

The rate of rhythmicity of smooth muscle, in general, is quite slow, varying from one contraction every five seconds up to one contraction every few minutes. This is in contrast to the rhythmic contraction of cardiac muscle more than once every second.

The rhythmic contraction of smooth muscle is responsible for many of the functions of smooth muscle organs. For instance, rhythmic contractions of the uterus at the end of pregnancy are responsible for birth of the baby, and the rhythmic contractions of the gut are responsible for peristalsis that propels food along the intestines.

Tone of smooth muscle. Almost all the smooth muscle in the body remains at least mildly contracted continuously even without stimulation by nerves, unlike skeletal muscle. This continuous contraction is called tone. Figure 211 illustrates the difference between rhythmic contractions and tonic contraction. Throughout the record the rhythmic contractions cause repetitive increases and decreases in muscle contraction. However, superimposed on these is a single prolonged tonic contraction.

Smooth muscle tone can vary greatly from time to time because of many factors such as fatigue of the muscle, availability of nutrients for muscular contraction, hormonal action, and irritation of the muscle.

Plasticity of smooth muscle. Another extremely important characteristic of smooth muscle is its ability to be stretched or shortened and still function equally well in both states. For instance, the urinary bladder when almost entirely empty can yet contract to a still smaller size and empty the remaining few milliliters of urine. On the other hand, after the bladder has been stretched to a volume of as much as one liter, the muscle fibers are still capable of contracting and forcing the urine to the exterior. In other words, smooth muscle is different from skeletal muscle in that its effectiveness of contraction is great at many lengths.

The veins offer another important example of the plasticity of smooth muscle. When the blood volume becomes great the veins simply stretch to accommodate most of the extra volume, and they do so without causing a great increase in venous pressure. When the blood volume finally becomes readjusted to normal, the veins assume

smaller volume, always maintaining almost constant pressure on the inside.

The principle of plasticity applies also to the accumulation of food in the gastrointestinal tract and to the function of the gallbladder, the bile ducts, and almost all hollow organs of the body. All these organs maintain essentially the same internal pressure despite great changes in their volumes.

Control of smooth muscle contraction. The most important factors controlling the rhythmic and tonic contractions of smooth muscle are (1) the amount of stretch applied to the muscle, (2) local irritation, and (3) the effects of the autonomic nervous system.

Ordinarily, stretch of a smooth muscle organ or local irritation of the organ increases the intensity of both the tonic and rhythmic contractions. Occasionally, though, very strong irritation causes the smooth muscle to go into such an intense state of tonic spasm that the rhythmic contractions are no longer apparent. This is especially characteristic of the gastrointestinal tract when it is irritated by an infectious disease, by an irritating food or drug, or by trauma.

CONTROL OF SMOOTH MUSCLE BY THE AUTONOMIC NERVOUS SYSTEM. The autonomic nervous system, which will be discussed in Chapter 26, is composed of two separate parts, the *sympathetic* and the *parasympathetic systems*. Nerve fibers pass from each of these to most smooth muscle of the body. When stimulated, the sympathetics secrete a hormone called *norepinephrine* and smaller amounts of a similar hormone, *epinephrine,* and the parasympathetics secrete a hormone called *acetylcholine.* Norepinephrine and epinephrine stimulate the smooth muscle in some organs, but in other organs inhibit it. Acetylcholine, on the other hand, inhibits all smooth muscle that the other two hormones stimulate and stimulates all that the other two inhibit. In other words, the hormones secreted by the two portions of the autonomic nervous system have exactly opposing effects. Some examples of this are the following: (1) The pupil of the eye is constricted by the parasympathetics but dilated by the sympathetics. (2) The coronary arteries are dilated by the sympathetics but constricted by the parasympathetics. (3) The intestines are constricted by the parasympathetics but dilated by the sympathetics. (4) The sphincters of the gut are constricted by the sympathetics but dilated by the parasympathetics. (5) The sphincters that regulate emptying of the urinary bladder are constricted by the sympathetics but dilated by the parasympathetics.

The reason why norepinephrine contracts some but relaxes other smooth muscle, while acetylcholine causes exactly the opposite effects, is not known, but this is believed to result from a difference in "receptor substances" in the muscle cells. If the receptor substance is of an excitatory nature, norepinephrine will cause stimulation; if of an inhibitory nature, it will cause inhibition. It is presumed that the excitatory substance for norpinephrine is an inhibitory substance for acetylcholine, and the inhibitory substance for norepinephrine is an excitatory substance for acetylcholine. This would explain the antagonistic effects of the two hormones on smooth muscle.

REFERENCES

Bourne, G. H., editor: *The Structure and Function of Muscle.* 3 Volumes. New York, Academic Press, 1960.

Bowen, W. P.: *Applied Anatomy and Kinesiology.* 7th Ed. Philadelphia, Lea & Febiger, 1953.

Gelfan, S.: Muscle. *Ann. Rev. Physiol.* 20:67, 1958.

Gerard, R. W., and Taylor, R. E.: Muscle and nerve—physiologic orientation. *Am. J. Med.* 15:83, 1953.

Hill, A. V.: The heat production of muscle and nerve, 1848–1914. *Ann. Rev. Physiol.* 21:1, 1959.

Hill, D. K.: Muscle. *Ann. Rev. Physiol.* 13:99, 1951.

Mommaerts, W. F., Brady, A. J., and Abbott, B. C.: Major problems in muscle physiology. *Ann. Rev. Physiol.* 23:529, 1961.

Nachmansohn, D., editor: *Molecular Biology. (Elementary Processes of Nerve Conduction and Muscle Contraction.)* New York, Academic Press, 1960.

Ralston, H. J.: Mechanics of voluntary muscle. *Am. J. Phys. Med.* 32:166, 1953.

Tiegs, O. W.: Innervation of voluntary muscle. *Physiol. Rev.* 33:90, 1953.

SYNAPTIC FUNCTIONS OF NEURONS, AND DESIGN OF THE NERVOUS SYSTEM

22

Basic Organization of the Nervous System

The basic functional units of the nervous system are the *neurons,* which transmit information from one part of the nervous system to another by way of *nerve fibers* that are outgrowths of the neurons. Also, neurons in the central nervous system differentiate the incoming information from the periphery and determine the impulses to be transmitted back to the different parts of the body to perform the various bodily activities.

The central nervous system is composed of three major parts: (1) the *sensory system* which transmits information from the body to the central nervous system, (2) the *motor system* which transmits signals from the central nervous system back to the active parts of the body, and (3) the *integrative system* which assesses information, stores information, and determines what signals to transmit either elsewhere in the central nervous system or into the motor nerves.

The reflex arc. Many functions of the nervous system are performed by *simple reflexes* which are accomplished by three basic nervous units: a receptor, a transmitter, and an effector. A *receptor* is any type of sensory nerve ending that is capable of detecting one or more of the usual bodily sen-

sations, such as touch, pressure, smell, and sight. Once the sensation is detected, an impulse is transmitted by the *transmitter,* which may be either a single neuron or several successive neurons that connect in series with each other. The *effector* may be a skeletal muscle or one of the internal organs such as the heart, gut, glands, and so forth that can be controlled by nerves.

The simplest reflex arc is the so-called *axon reflex* which will be described in Chapter 25. In this, the reflex is transmitted over a single neuron; a sensory receptor in the skin is stimulated by a scratch or any other noxious stimulus, and impulses are transmitted to other branches of the same sensory fiber, the endings of which emit a vasodilator substance that causes the blood vessels to dilate.

The next level of reflex arc involves the spinal cord, in which case sensory signals are transmitted into the cord, and these are returned directly back to the muscles or other effectors. For instance, the so-called *stretch reflex* operates in this manner; if a muscle is suddenly stretched, a sensory impulse from the muscle is transmitted to the spinal cord by one neuron, and a second neuron transmits an impulse directly back to the muscle, causing it to tighten and thereby resist the stretch. This is called a *two neuron reflex.*

Multiple neuron reflexes are also abundant in the spinal cord. For instance, if a person places his hand on a hot stove, pain impulses travel into the gray matter of the cord where they excite a series of neurons before impulses are transmitted back to the muscles. The neurons are so organized that they do not simply send an instantaneous barrage of signals back to the muscles but, instead, send specially selected signals that cause those muscles to contract that will lift the hand away from the painful object. This multineuronal reflex is called a *withdrawal reflex.*

Other reflex arcs utilize the brain as well as the cord and peripheral nerves. For instance, if a person is struck in the face by an adversary, sensory signals are transmitted all the way into the brain, where the person weighs the evidence prior to reacting. If the adversary is beyond a certain critical size, the person runs, but if conditions are more propitious, he strikes back. Even though considerable time, thought, memory, and judgment enter into this reaction, it is nothing more than a very complicated reflex.

Therefore, in essence, the nervous system operates by a multitude of different types of reflexes. These reflexes may be either very simple ones that always occur in exactly the same manner time after time or very complex ones that are modified by previous memory, by assessment of information from many different sources, and by judgment of the best course of activity. Indeed, even memories are originally initiated by sensory input signals; therefore, some acts that we perform today are the reflex effects of sensory input signals that entered the brain many years ago.

FUNCTION OF THE SYNAPSE

The synapse is the juncture between two neurons. It is through the synapse that signals are transmitted from one neuron to another. However, the synapse has the capability of transmitting some signals and refusing other signals, thereby making it a valuable tool of the nervous system for choosing which course of events to follow. It is because of this variable transmission of signals that the synapse itself is perhaps the most important single determinant of central nervous system function.

Physiologic Anatomy of the Synapse

Figure 212 illustrates a neuron, which is composed of three major parts: the *soma* which is the main body of the neuron and two types of projections, the *dendrites,* which, except in the case of sensory neurons, are usually short projections that extend only a few millimeters in the nervous system, and the *axon* which may extend as far as several feet within the central nervous system or out into peripheral nerves. The figure also shows hundreds of small fibers leading to the neuron and terminating in small knobs called *presynaptic terminals.* These lie on the surfaces of the soma

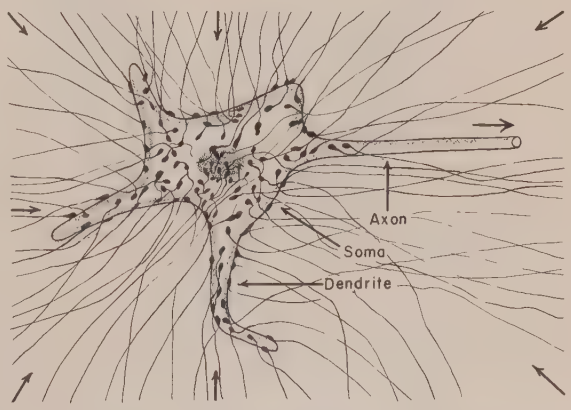


Figure 212. A typical neuron, showing hundreds of presynaptic terminals that originate from other neurons.

and dendrites. It is this juncture between the presynaptic terminals and dendrites or soma that is called the *synapse.* The small fibers are many branches of axons that come from other neurons.

Transmission of the Impulse at the Synapse

Transmission of the impulse from the presynaptic terminals to the dendrites or soma of the neuron occurs in very much the same way that transmission occurs at the neuromyal junction as explained in Chapter 20. However, there is one major difference: at the neuromyal junction the acetylcholine secreted by the nerve endings can only excite the muscle fiber, while at the synapse some presynaptic terminals secrete an *excitatory transmitter substance* and others secrete an *inhibitory transmitter substance.*

Excitation of the neuron—"the excitatory transmitter." Figure 213 illustrates the structure of a typical presynaptic terminal lying adjacent to the membrane of either the soma or dendrite of a neuron. The terminal contains many small globules of transmitter substance, and when a nerve impulse reaches the presynaptic terminal, momentary changes in the membrane structure of the terminal allow a few of these globules to escape into the synaptic space shown in the figure. The transmitter substance then acts on the surface of the neuron to cause either excitation or inhibition.

The exact chemical nature of neither the excitatory nor inhibitory substance has yet been proved. However, it is believed that the excitatory transmitter, at least in part of the central nervous system, is *acetylcholine.* There are two reasons for this belief. First, in the ganglia of the sympathetic chain, large quantities of acetylcholine are secreted during the transmission of signals through the synapses. Second, certain "cholinergic" nerve fibers are known to enter the substance of the spinal cord from the exterior and to excite neurons inside the spinal cord. Since it is known that these neurons secrete acetylcholine, it is presumed that acetylcholine is the excitatory transmitter.

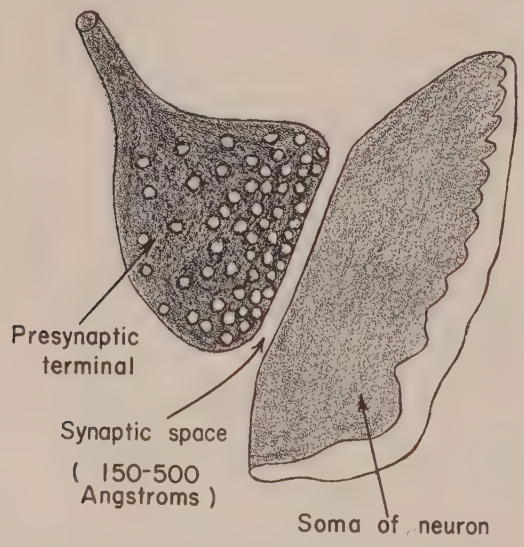


Figure 213. Physiologic anatomy of the synapse.

Excitatory postsynaptic potential. The manner in which the excitatory transmitter excites the neuron can be explained by reference to Figure 214. The excitatory transmitter increases the permeability of the neuronal membrane immediately beneath the presynaptic terminal. This allows sodium ions to flow rapidly to the inside of the cell, and since sodium ions carry positive charges, the net result is an increase in positive charges inside the cell. Therefore, an electrical potential, illustrated by the arrows in Figure 214, is immediately set up around the soma of the cell, called the *excitatory postsynaptic potential.* If this potential becomes great enough, it will initiate an action potential in the axon that will travel over the nerve fiber leading from the neuron. However, if the excitatory postsynaptic potential is less than a certain threshold value, nothing will happen—that is, no action potential.

Summation of excitatory postsynaptic potentials. Almost invariably, stimulation of a single presynaptic terminal will not initiate an impulse in the axon. Instead, large numbers of presynaptic terminals must become excited at the same time. Referring once again to Figure 212, it is readily apparent that hundreds of presynaptic terminals could easily be excited simultaneously and that these operating in unison could cause neuronal discharge.

That is, if two terminals release their excitatory transmitter substances simultaneously, twice as much sodium enters the cell body and twice as much excitatory postsynaptic potential develops. This is called *summation.* As more and more terminals summate, the potential becomes greater and greater, until finally it is great enough to excite the axon.

SPATIAL AND TEMPORAL SUMMATION. Two different types of summation occur at the synapse in the same way that two types of summation occur in nerve and muscle fibers. These are spatial and temporal summation. *Spatial summation* means that two or more presynaptic terminals fire simultaneously, thereby "summing" their individual effects on the excitatory postsynaptic

potential. *Temporal summation* means that the same presynaptic terminal fires two or more times in rapid succession, thus adding the effect of the second firing to that of the first. The effect of the first discharge usually lasts for about 15 milliseconds. Therefore, if two successive discharges of the same presynaptic terminal occur within less than 1/70 second of each other, temporal summation occurs, and the closer together these discharges, the greater will be the degree of summation.

Repetitive discharge of the axon—threshold for firing. Once sufficient excitatory transmitter has been secreted by the presynaptic terminals to raise the postsynaptic potential above a critical value called the *threshold* of the neuron, the axon will fire repetitively and will continue to do so as long as the potential remains above this threshold. Furthermore, as the postsynaptic potential rises higher and higher above the threshold, the more rapidly will the axon fire. For instance, let us assume that an excitatory postsynaptic potential of 15 millivolts is the threshold value. If the potential rises slightly above this, to 16 millivolts, we would expect the axon to discharge perhaps 5 to 10 times per second. But, if the postsynaptic potential rises to as high as 30 millivolts, twice the threshold value, then the axon perhaps would fire as many as 200 to 300 times per second.

FACILITATION AT THE SYNAPSE. When presynaptic terminals discharge but fail to cause an action potential in the axon, the neuron still becomes *facilitated*; that is, even though an action potential does not result, the neuron becomes more excitable to impulses from other presynaptic terminals. Let us assume, for instance, that 25 presynaptic terminals must fire simultaneously to discharge a neuron. If 20 presynaptic terminals fire simultaneously, the neuron will not discharge, but it does become sufficiently "facilitated" so that any 5 additional presynaptic terminals anywhere on the surface of the cell could now elicit an impulse. This is the situation that develops in a "nervous" person, for large numbers of his neurons become facilitated though not excited; yet very few extraneous impulses can then cause terrific reactions.

RESPONSE CHARACTERISTICS OF DIFFERENT NEURONS. The nervous system is composed of hundreds of different types of neurons. Some of the cell bodies are tremendously large in size while others are very minute. Some are capable of transmitting as many as 1000 impulses per second over their axons and some are capable of transmitting no more than 25 to 50 per second. And, different neurons have different thresholds. All of these differences are fortunate, because neurons in different parts of the brain perform different functions.

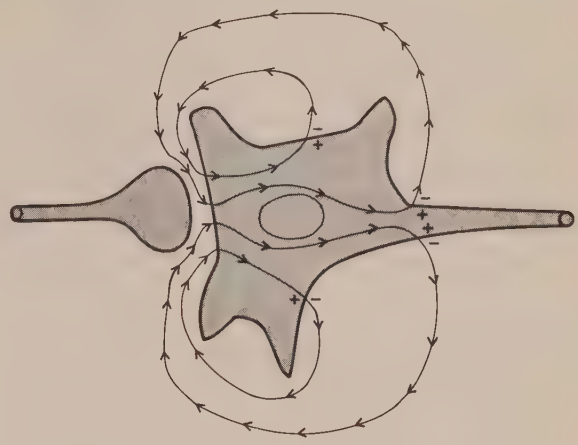


Figure 214. Flow of current beneath the synapse and around the neuron, showing that a synaptic discharge causes increased current flow over the entire body of the neuron.

Figure 215 illustrates typical *response patterns* of three different neurons. Neuron 1 requires a postsynaptic potential of 5 millivolts to discharge it, and it reaches a maximum discharge rate of 100 per second. The threshold of neuron 2 is 8 millivolts, and the maximum frequency of discharge is 35. Finally, neuron 3 has a high threshold of 18 millivolts, but once it becomes excited, its maximum discharge rate is 120 per second.

Inhibition at the synapse. Some of the presynaptic terminals secrete an *inhibitory transmitter* instead of an excitatory transmitter. Unfortunately, the inhibitory transmitter, like the excitatory transmitter, has not been isolated with certainty. However, a substance called *gamma aminobutyric acid* (GABA), found in large quanti-

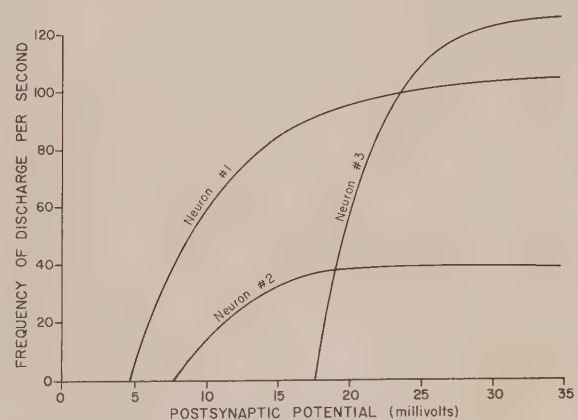


Figure 215. Response characteristics of different types of neurons.

ties in inhibitory nerve fibers, has an inhibitory effect on the synapse. Therefore, it is now believed to be the inhibitory transmitter.

The inhibitory transmitter has an opposite effect on the synapse to that caused by the excitatory transmitter, actually causing a negative potential called the *inhibitory postsynaptic potential.* An example of the manner in which the inhibitory transmitter operates at the synapse is the following: Let us assume that the threshold for stimulation of a neuron is +10 millivolts and that excitatory presynaptic terminals are generating a postsynaptic potential of +20 millivolts. The neuron will be firing repetitively at a rate of perhaps 50 to 200 times per second. At this time, stimulation of inhibitory presynaptic terminals releases inhibitory transmitter which we will assume creates an inhibitory postsynaptic potential of −15 millivolts. When this summates with the excitatory potential of +20 millivolts, the net value is only +5 millivolts, a level far below the threshold of the neuron. Therefore the neuron stops discharging.

EXCITATORY AND INHIBITORY NEURONS. The central nervous system is made up of *excitatory neurons* that secrete excitatory transmitter at their nerve endings, and of *inhibitory neurons* that secrete inhibitory transmitter at their nerve endings. Certain neuronal centers of the central nervous system are entirely excitatory while others are both excitatory and inhibitory, comprised of both types of cells. Therefore, the central nervous system, unlike the peripheral sensory and motor systems, has two modes of activity, either excitation or inhibition instead of simply excitation alone.

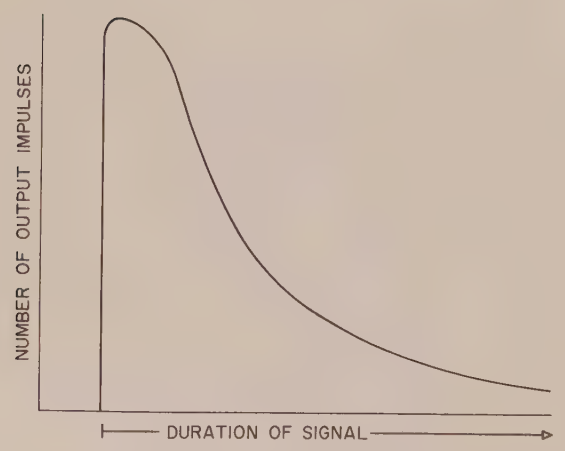


Figure 216. Effect of fatigue on the number of output impulses from the neuron after it begins to be stimulated.

Some Special Characteristics of Synaptic Transmission

One-way conduction at the synapse. Impulses traveling over the soma or dendrites of the neuron cannot be transmitted backward through the synapses into the presynaptic terminals. Thus, only one-way conduction occurs at the synapse. This is extremely important to the function of the nervous system, for it allows impulses to be channeled in the desired direction.

Fatigue of the synapse. Transmission of impulses at the synapse is different from transmission in nerve fibers in a very important respect: the synapse *fatigues* very rapidly while nerve fibers fatigue hardly at all. Figure 216 illustrates this effect, showing that at the beginning of the input signal, the neuron discharges very rapidly at first but then more and more slowly the longer it is stimulated. Some synapses fatigue very rapidly, while others fatigue very slowly.

One might expect this phenomenon of fatigue to be an impediment to the action of the central nervous system, but, on the contrary, it is a necessary feature. Were it not for synaptic fatigue, a person could never stop a thought, a rhythmic muscular activity, or any other prolonged repetitive activity of the nervous system once it had begun. Fatigue of synapses is a means by which the central nervous system allows a nervous reaction to fade away to make way for others. Later in the chapter, specific instances of the importance of fatigue are pointed out.

The "memory" function of the synapse. When large numbers of impulses pass through a synapse, the synapse becomes "permanently" facilitated so that impulses from the same origin can pass through the synapse with greater ease at a later time. It is believed that this is the means by which memory occurs in the central nervous system. For instance, a given thought initiated by visual, auditory, or any other type of signal causes impulse transmission through given pathways in the brain. If the same thought is repeated over and over again, such as the thought caused by seeing the same view again and again, then the pathways for the thought become permanently facilitated so that subsequent impulses pass through these pathways with the greatest of ease. Therefore, impulses may later enter the same pathways from some other source besides the visual apparatus, and the person "remembers" the scene rather than actually seeing it. This process of memory will be discussed in greater detail in Chapter 24.

Other factors that affect synaptic transmission. Figure 217 illustrates several different factors that

can alter the number of impulses transmitted by a synapse in response to a given degree of stimulation of the presynaptic terminals. Note that *hypnotics, anesthetics,* and *acidosis* all depress the transmission of impulses at the synapse, while *alkalosis,* the *mental stimulants* such as *caffeine, benzedrine,* and *strychnine* all greatly facilitate synaptic transmission. Strychnine, when given in sufficiently high dosage, will cause the neurons to discharge spontaneously even in the absence of a presynaptic stimulus. It is in this way that strychnine kills an animal, causing so many neuronal impulses to be transmitted throughout the central nervous system and into the motor system that it causes death by spastic paralysis of the respiratory muscles.

The integrative function of the neuron. To summarize the action of the neuron, one can call it an "integrator," which means a type of calculator that collects information and sums it all together. Signals reach the neuron by way of excitatory and inhibitory presynaptic terminals that in turn are excited by neurons from other parts of the nervous system. If the resultant sum of all the excitatory and inhibitory effects is above the threshold for excitation, the neuron fires. Some of the neurons fatigue rapidly, others slowly. Some of the neurons have high thresholds, others have low thresholds. Some fire at rapid rates, others at slow rates. The varying characteristics of the different neurons and their different connections in the nervous system allow one portion of the nervous system to control one function of the body while another portion controls another function. They allow the sorting of signals to determine their meanings, the performance of special skilled motions, the thinking of specific thoughts, and modification of these thoughts by signals arriving from other parts of the nervous system.

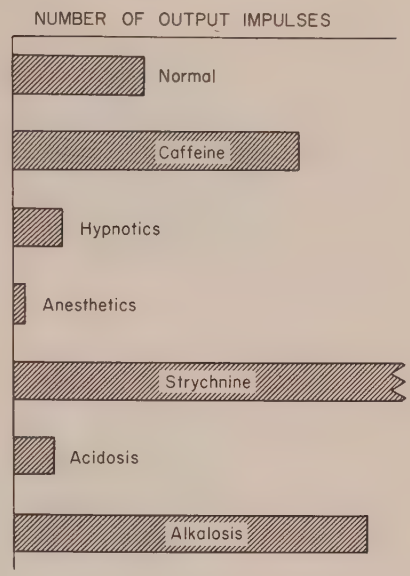


Figure 217. Effect of different drugs and different physiological states on the excitability of a neuron.

BASIC NEURONAL CIRCUITS

The Neuronal Pool

The central nervous system is divided into many different anatomic parts, in each of which are located many accumulations of neurons called *neuronal pools.* An important feature of the different neuronal pools is that each has a different pattern of organization from all the others. That is, the distribution of nerve fibers within the pool, the number of incoming nerve fibers, the number of outgoing fibers, the types of neurons in the pool, and many other features differ from one pool to another. Each pool is organized to perform a specific function. The purpose of the present section is to discuss, first, the general functions of the neuronal pool and, second, the characteristics of special types of pools.

Area of stimulation. Figure 218A shows a single nerve fiber entering a neuronal pool. This fiber branches in many directions, ending in as many as 100 different terminal branches that spread within the pool. Often several of the terminals end on the same neuron, while others spread out singly to many nearby neurons.

Figure 218B shows the sphere of influence of the incoming nerve fiber. In the very center is an area where a sufficient number of terminals end on each neuron to cause excitation every time the incoming fiber is stimulated. This is called the *discharge zone.* In other words, each time the nerve fiber is stimulated, the neurons in this area also become stimulated and they in turn transmit signals out of the pool.

Facilitation in the neuronal pool. Figure 218B also shows a *facilitated zone* surrounding the discharge zone. The neurons in this area receive terminals from the nerve fiber but not enough to cause excitation. Yet the firing of the terminals does make these cells more excitable than usual, so that impulses arriving in the pool from other incoming nerve fibers can stimulate them far more easily than usual.

Simple Circuits in Neuronal Pools

The simplest circuit in a neuronal pool is that in which one incoming nerve fiber stimulates one

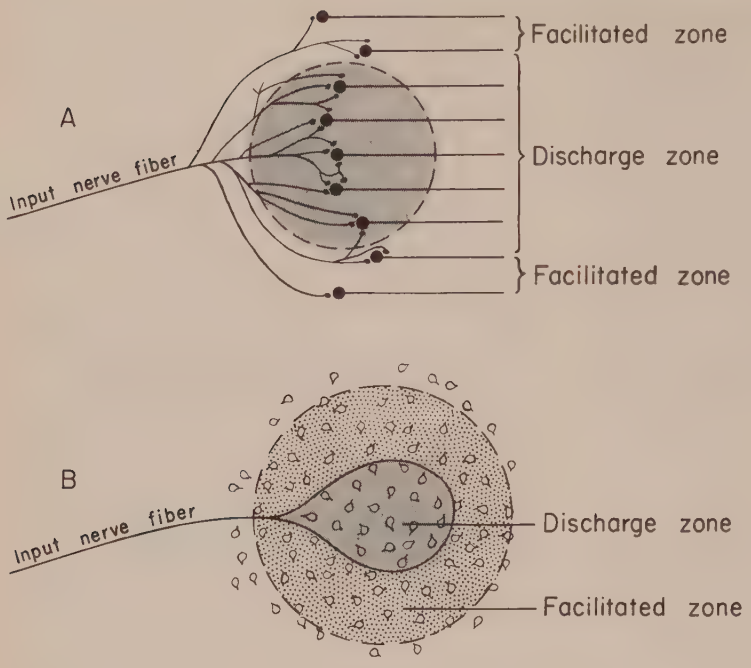


Figure 218. (A) Typical organization of a neuronal pool. (B) "Discharge" and "facilitated" zones of a neuronal pool.

outgoing fiber. This type of circuit does not exist in a precise form, though occasionally it is approximated. For instance, in transmitting certain types of sensory signals from peripheral nerves into the brain, a single incoming impulse into a neuronal pool may cause a single outgoing impulse; more than likely, though, there will be some simple ratio such as a 1 to 2, or 2 to 1, or so forth.

Such a circuit obviously acts simply as a relay station, relaying on to additional neurons essentially the same information that enters it.

The diverging or amplifying circuit. Figure 219A illustrates a *diverging circuit,* showing a single incoming nerve fiber entering the neuronal pool from above. This fiber branches to stimulate two neurons, the fibers from these branch to stimulate still more neurons, and the branching continues until the desired number of output impulses is achieved. This is also an *amplifying* type of circuit, one incoming impulse causing many output impulses.

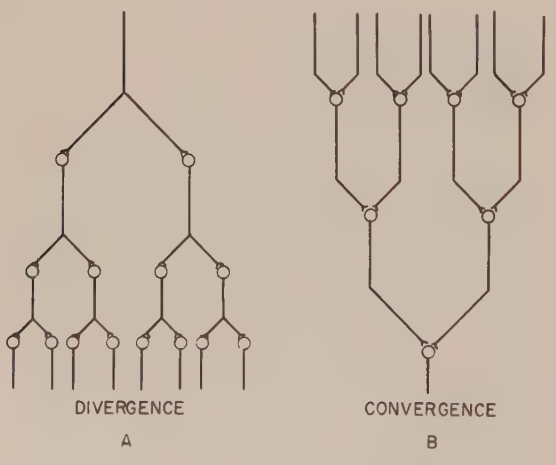


Figure 219. (A) The diverging circuit. (B) The converging circuit.

The diverging type of circuit is especially exemplified by the control of the skeletal muscles. Stimulation of a single *motor cell* in the brain sends an impulse down to a neuronal pool of *internuncial cells* in the spinal cord. Here the impulse is amplified so that 50 to 1000 impulses may be transmitted anteriorly in the spinal cord to stimulate many *anterior horn cells.* Fibers from each of these then pass to the muscles where each one of them branches to stimulate approximately 150 muscle fibers. Thus, a single motor neuron of the brain stimulates as many as 150,000 muscle fibers.

The converging circuit. Figure 219B illustrates a *converging circuit,* which is the opposite of the diverging circuit. In this instance, impulses enter the neuronal pool from eight different nerve fibers but stimulate only four neurons. These then stimulate two neurons which in turn stimulate only a single neuron. This type of circuit allows impulses from many different sources to cause the same reaction. It is especially characteristic of the sensory nervous system. For instance, impulses coming from the eyes, ears, skin, and many other sources can cause a person to think the same thought. The smell of a pigpen might give a person

the idea of a pig, the sight of a pig could do the same, the sound of his grunting could also produce this idea, and touching the pig or eating pork chops might lead to the same idea. In other words, sensations from many sources converge on the same part of the brain to give the same final thought.

USE OF CONVERGING CIRCUITS TO PERFORM INTEGRATIVE FUNCTIONS. Converging circuits can provide integrative functions that are much more complex than the integrative function of single neurons, for in a converging circuit there may be cells that respond to the incoming signal in several different ways at the same time. For instance, one of the neurons might have a very high threshold but when once stimulated might be so powerful that it causes a very intense reaction. In other parts of the circuit there may be neurons that have low thresholds but when stimulated transmit only weak signals. The entire pool, therefore, could transmit very weak output signals usually but very, very powerful signals if a certain degree of input stimulation should be exceeded. Thus, the converging circuit can be organized to perform almost any type of selective function. The individual neurons of the circuit select which signals shall pass, but the arrangement of the fibers and the combinations of cells determine which incoming signals will have the greater effects, and whether the effects will be excitatory or inhibitory.

Repetitive circuits. THE REVERBERATING CIRCUIT. Figure 220 shows two different types of circuits which, when stimulated only once, will cause the output cell to transmit a series of impulses. The upper circuit of this figure is a *reverberating circuit* which functions as follows: An incoming signal stimulates the first neuron which then stimulates the second and third cells. However, branches return to the first neuron and restimulate it. As a result, the signal travels once again through the chain of neurons, this process continuing around and around the circuit indefinitely until one or more of the cells fails to fire. The usual cause of failure is fatigue, and until this takes place, the output neuron continues to be stimulated every time the signal goes around the circuit, giving a continuous output signal.

The reverberating circuit is the basis of innumerable central nervous system activities, for it allows a single input signal to elicit a response lasting a few seconds, minutes, or hours. Indeed, the life-long respiratory rhythm is caused by a reverberating circuit that continues to reverberate without fatiguing as long as the person remains alive. It is probable, also, that a special circuit in the brain causes a person to awaken when it reverberates and allows him to relax into a state of sleep when it stops reverberating. Almost all rhythmic muscular activities, including the rhythmic movements of walking, are probably controlled by reverberating circuits.

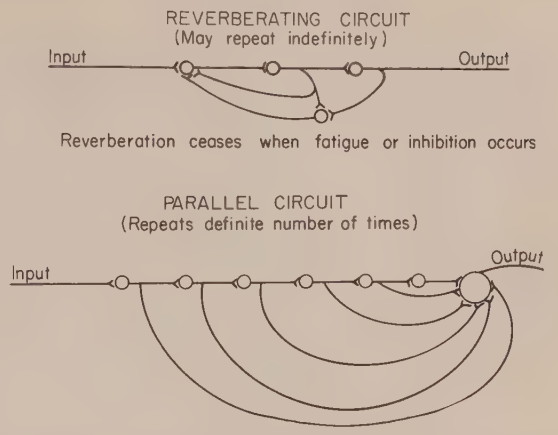


Figure 220. Reverberating and parallel circuits for repetitive discharge.

A few moments of thought about the reverberating type of circuit will emphasize the extreme number of variable functions it can perform. For instance, the number of neurons in the circuit may be great or small. If the number is great, the length of time required for the signal to go around the circuit will be long; if the number is small, the reverberatory period will be short. As a result, the output signal may repeat itself rapidly or slowly. Furthermore, more than one of the neurons in the circuit can give off output signals. One of the cells in the circuit, for instance, might cause an arm to move upward, then another cell in the circuit a fraction of a second later might cause the arm to move to the right, another cell still later cause it to move downward, and another cell cause it to move to the left. As the reverberatory cycle repeats itself once more, the same motions occur again, causing the arm to move around and around continuously until the cycle stops. When the person wishes to perform this motion, all he has to do is to set off the reverberatory cycle.

THE PARALLEL CIRCUIT. The lower part of Figure 220 illustrates the *parallel* type of repetitive circuit in which a single input signal stimulates a sequence of neurons that send separate nerve fibers directly to a common output cell. Because a delay of about ½000 second occurs each time an impulse crosses a synapse, the impulse from the first neuron arrives at the output cell ½000 second ahead of the impulse from the second neuron, and impulses continue to arrive at these short intervals until all the neurons have been stimulated. Since

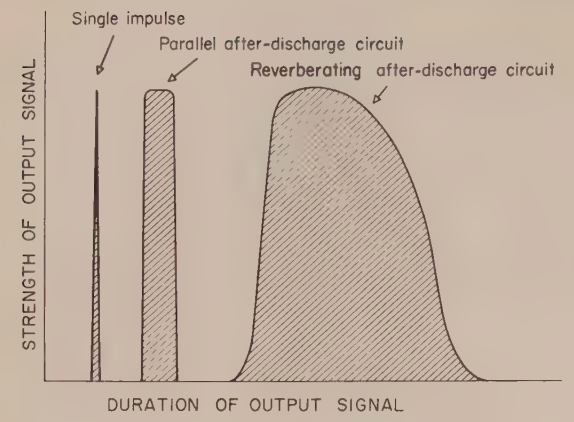


Figure 221. Characteristics of signals caused by (1) a single nerve impulse, (2) discharge of a parallel after-discharge circuit, and (3) discharge of a reverberatory after-discharge circuit.

there is no feedback mechanism in the circuit, the repetition then ceases entirely.

Parallel circuits in different parts of the central nervous system differ in several ways: First, some circuits are composed of far more neurons than others and give a repetitive output lasting considerably longer periods of time. Second, the parallel circuit can be combined with diverging circuits, so that the signal can be amplified to stimulate the output neuron very strongly during part of the repetitive discharge and weakly during other parts of the discharge. In this way the output signal will have a definite amplitude pattern as well as a definite duration.

The parallel type of circuit has certain advantages over the reverberating circuit. For instance, the very variable phenomenon of fatigue is a major factor in determining how long the reverberating circuit will continue to fire, whereas the output duration of the parallel circuit is independent of fatigue. For performing very exact activities such as mathematical calculations the parallel circuit is perhaps used more than the reverberating circuit, while for rhythmic functions or for greatly prolonged discharges the reverberating circuit is probably the one used. Unfortunately, the parallel circuit cannot control functions that last more than a fraction of a second, because one neuron is required in the circuit for each ½₀₀₀ second. Probably not more than a few hundred neurons are ever organized into parallel circuits, which would limit this type of circuit to activities occurring in less than about ¹⁄₁₀ second.

Figure 221 illustrates the time duration and amplitude characteristics of the output signals from, first, a one-to-one circuit, second, a parallel after-discharge circuit, and, third, a reverberatory after-discharge circuit.

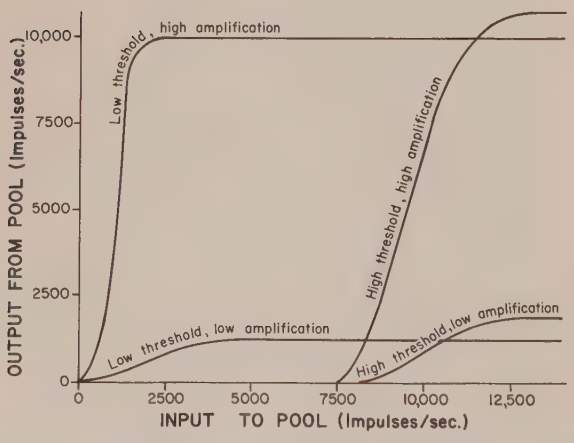


Figure 222. Function curves of different neuronal pools, illustrating the response of different types of circuits to incoming impulses.

Relationship of the Output Signal to Input Signal of a Neuronal Pool

Figure 222 illustrates some of the response characteristics of typical neuronal pools. The four different types of responses illustrated here are: (1) low threshold, high amplification response, (2) low threshold, low amplification response, (3) high threshold, high amplification response, and (4) high threshold, low amplification response.

With a little imagination one can readily understand possible functions for each of these types of circuits. For instance, the *low threshold, high amplification* response would be that in which a person reacts vigorously to a very slight sensory input signal, such as stimulation of only a few pain nerve endings. The *high threshold, high amplification* response would be one in which a person subjected to moderately loud noises pays no attention and yet when a very loud noise suddenly occurs, he reacts vigorously. In the same manner, we could easily find typical central nervous system responses representative of the other two response curves in Figure 222.

An important characteristic of neuronal pools is that the response curve to input signals from one source can be altered by input signals coming from secondary sources. For instance, sensory nerve impulses from the skin into the spinal cord

ordinarily cause no reflex skeletal muscle effects. However, strong facilitatory impulses transmitted from the brain down the spinal cord to the neurons that control the muscles can make these neurons so excitable that a very light scratch on the skin elicits a strong contraction of the underlying muscle. Thus, a high threshold, low amplification circuit has been changed into a low threshold, high amplification circuit.

It is now up to the imagination of the student to conceive the many possible ways in which individual types of neurons can be organized into different types of neuronal pools and in which different types of neuronal pools can perform an infinite number of reflex and integrative nervous functions. Unfortunately, the precise circuits in most neuronal pools have not yet been worked out in detail. Nevertheless, the types of circuits that are already known can suggest mechanisms by which all of the functions of the central nervous system could be attained.

REFERENCES

Bishop, G. H.: Natural history of the nerve impulse. *Physiol. Rev.* 36:376, 1956.

Eccles, J. C.: Neuron physiology—introduction. *Handbook of Physiology* 1:59, 1959.

Eccles, J. C.: *The Physiology of Nerve Cells.* Baltimore, Johns Hopkins University Press, 1957.

Elliott, K. A. C., and Jasper, H. H.: Gamma-aminobutyric acid. *Physiol. Rev.* 39:383, 1959.

Fadiga, E., and Brookhart, J. M.: Monosynaptic activation of different portions of the motor neuron membrane. *Am. J. Physiol.* 198:693, 1960.

Fernandez-Moran, H., and Brown, R.: *Submicroscopic Organization and Function of Nerve Cells.* New York, Academic Press, 1958.

Florey, Ernst: Comparative physiology: transmitter substances. *Ann. Rev. Physiol.* 23:501, 1961.

Furshpan, E. J.: Neuromuscular transmission in invertebrates. *Handbook of Physiology* 1:239, 1959.

Grundfest, H.: Synaptic and ephaptic transmission. *Handbook of Physiology* 1:147, 1959.

Ortmann, R.: Neurosecretion. *Handbook of Physiology* 2:1039, 1960.

Spyropoulos, C. S., and Tasaki, I.: Nerve excitation and synaptic transmission. *Ann. Rev. Physiol.* 22:407, 1960.

SOMESTHETIC SENSATIONS AND INTERPRETATION OF SENSATIONS BY THE BRAIN

23

SOMESTHETIC SENSATIONS

The term *somesthetic sensation* means sensations from the body; these include *exteroceptive* and *proprioceptive* sensations. Exteroceptive sensations are those normally felt from the skin, such as touch, pressure, heat, cold, and pain. Proprioceptive sensations are those that keep the brain apprised at all times of the physical state of the body, including such sensations as tension on the muscles, tension on the tendons, angulation of the joints, and deep pressure from the bottom of the feet. The sensations from the internal organs of the body, the *visceral sensations,* have many qualities very similar to those of the exteroceptive sensations. Therefore, the visceral sensations are considered to be part of the exteroceptive sensations.

Transmission of Sensations to the Brain

Figures 223 and 224 illustrate respectively the general neuronal pathways by which exterocep-

tive and proprioceptive sensations are transmitted from the body to the brain. The sensations are detected by special nerve endings in the skin, in the muscles, the tendons, or the deeper areas of the body, and these transmit impulses through the nerve trunks into the spinal cord. At this point the nerve fibers branch, some of the branches ending in the spinal cord itself to cause "cord reflexes," which will be discussed in Chapter 25, and the others extending upward to the brain. The sensory pathways to the brain terminate mainly in the thalamus and cerebral cortex, which are the conscious areas that control perception and interpretation of the sensations. In addition to these areas, though, some branches end in various parts of the medulla, the pons, the mesencephalon, and the cerebellum. These latter areas are responsible for most of the automatic reactions that occur in response to somesthetic sensations, including the readjusting of equilibrium when one falls off balance, catching one's breath when suddenly exposed to very cold water, or jerking one's head to the side when strongly stimulated on that side. These automatic responses will also be discussed in Chapter 25. The present chapter is concerned

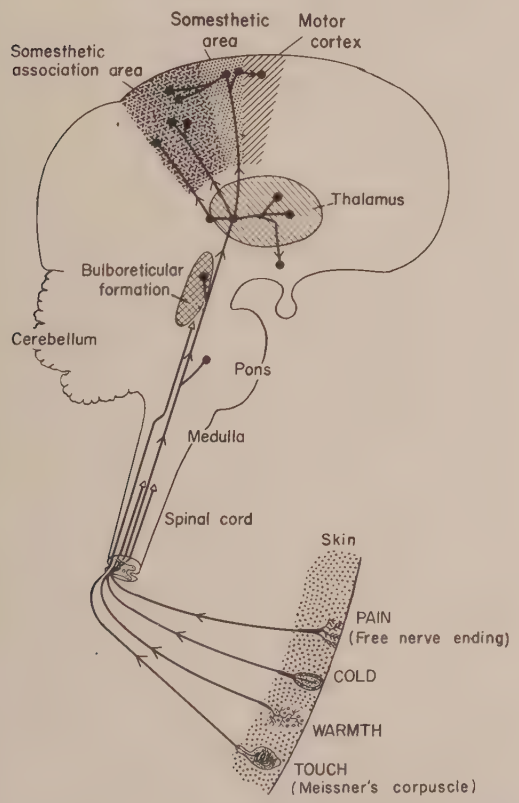


Figure 223. Transmission of exteroceptive sensations to the brain, showing the sensory receptors and the nerve pathways into the brain.

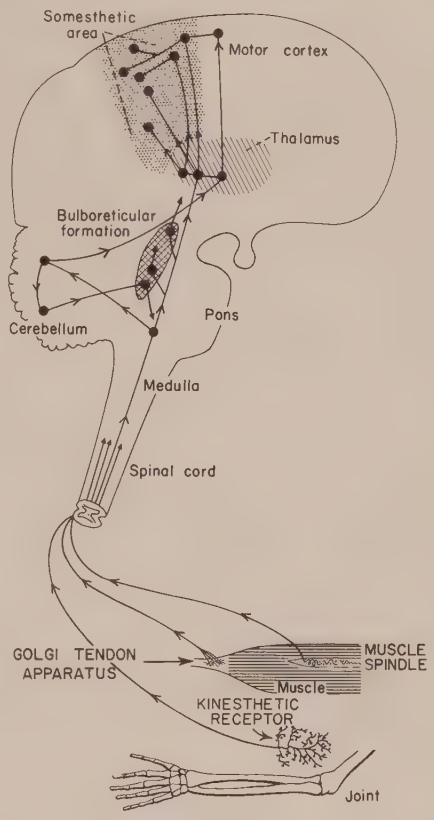


Figure 224. Transmission of proprioceptive sensations to the brain, showing the sensory receptors and the nerve pathways for transmitting these sensations into the brain.

principally with the pathways of sensation leading to voluntary reactions.

Close observation of the pathways shown in Figures 223 and 224 will demonstrate a few major differences between the exteroceptive and the proprioceptive pathways. Both of these types of sensations are transmitted to the posterior part of the thalamus and then relayed to other portions of the thalamus and to the cerebral cortex. But, in addition, the proprioceptive sensations are also transmitted about equally as strongly into the cerebellum, as shown in Figure 224. The cerebellum is the area that coordinates all the different movements of the body, but to do so, knowledge of the instantaneous position of all parts of the body must be available to the cerebellum at all times, which explains why such strong proprioceptive signals are transmitted into the cerebellum.

The Law of Specific Nerve Energies

The type of sensation that will be felt when a nerve fiber is stimulated is determined by the

point in the thalamus at which the fiber ends. One area of the thalamus gives rise to the sensation of pain, another the sensation of heat, another the sensation of cold, and so forth. Regardless of how a particular nerve fiber is stimulated, only one type of sensation will ever be felt. For instance, a nerve fiber that normally carries the sensation of heat can be stimulated by crushing it, stretching it, stimulating it electrically, placing acid on it, or otherwise, and in each instance the sensation will be that of heat. Likewise, a pain fiber stimulated in any way will give one the sensation of pain and of nothing else.

To state the law of specific nerve energies in another way, each type of sensation transmitted from the periphery to the brain is transmitted over a different set of nerve fibers.

The sensory receptors. Some sensory nerve endings are nothing more than small filamentous branches called *free nerve endings,* while others are special *end-organs* that are designed to respond only to special types of stimuli. Figures 223 and 224 illustrate some of the more representative sensory receptors, and their functions may be described as follows:

FREE NERVE ENDINGS. Free nerve endings detect the sensations of crude touch, pressure, pain, extreme heat, and extreme cold. These sensations are likely to become somewhat confused with each other because the free endings interconnect extensively. For instance, extreme heat or cold is likely to give one a sensation of pain, or very hard pressure might also be confused with pain.

Despite the fact that free nerve endings transmit only crude sensations, they are by far the most common type of nerve ending. They perform most of the general functions of sensation, while the specific functions, such as discrimination of very slight differences between degrees of touch, warmth, and cold, are left to the more specialized receptors.

SPECIAL EXTEROCEPTIVE RECEPTORS. Also shown in Figure 223 are specialized receptors for detection of cold, warmth, and light touch (Meissner's corpuscles). The reason why each of these end-organs registers only one type of sensation is not known. Presumably the physiologic organization of the end-organ itself allows cold, warmth, or light touch respectively to stimulate the nerve ending by some physical effect on the end-organ.

Still other specialized sensory receptors are present in the skin. For instance, at the base of each hair is a receptor that allows one to feel even the slightest pressure on any hair. In animals some of these hair receptors are so well developed that the associated hairs are called *tactile hairs.* A cat actually helps to guide itself in the darkness by its tactile whiskers. Special end-organs have also been found in the lips and snouts of some rooting animals; these presumably aid in the search for food. Finally, in the sexual organs special endings have been noted that are possibly responsible for the distinctive qualities of sexual sensations.

THE PROPRIOCEPTOR END-ORGANS. Figure 224 illustrates three different types of proprioceptor end-organs. One of them, the joint kinesthetic receptor, is found in the capsules of joints. These receptors apprise one of the *degree* of angulation of the joints and of the *rate* at which this degree of angulation changes.

Two special end-organs transmit proprioceptive information from the muscles. These are the *muscle spindle* and the *Golgi tendon apparatus.* The muscle spindle is composed of nerve filaments wrapped around specially adapted muscle fibers. The spindle detects the degree of stretch of the muscle. This information is transmitted to the central nervous system to aid in the control of muscular movements. The Golgi tendon apparatus detects the overall tension applied to the tendon, and, therefore, apprises the central nervous system of the effective strength of contraction of the muscle.

Adaptation of the sensory receptors. When a stimulus is suddenly applied to a sensory receptor, it usually responds very vigorously at first, but progressively less and less so during the next few seconds or minutes. An example of this is getting into a tub of hot water, which causes an intense burning sensation at first but after a moment's exposure produces a sensation of pleasing warmth. The same is true to varying degrees for all the other sensations. This loss of sensation during prolonged stimulation is called *adaptation* of the sensory receptor.

Adaptation occurs to much greater degrees for some sensations than for others. For instance, the sensations of light touch and of pressure adapt within a few seconds. This allows one to feel an object when he first touches it but to lose the sensation very soon thereafter. Were it not for this rapid adaptation of the light touch receptors, one would feel intense touch sensations, never stopping, from all the body areas in contact with any object such as the seat of a chair, the shoes, and even the clothing. These sensations would continue to bombard the brain to such an extent that one could hardly think of anything else. The value of rapid adaptation of certain of the sensations therefore is obvious.

The sensations of pain and proprioception usually adapt to only a slight extent. Pain sensations are elicited when tissue damage is occurring.

As long as the damage continues to occur it is important that the person also continue to be apprised of this fact, so that he will institute appropriate measures to remove the cause of the damage. The long persistence of proprioceptive sensations, also, is desirable because the brain needs to know the physical status of the different parts of the body at all times and not simply immediately after movements have occurred.

Modality of sensation—its determination by the thalamus

The term *modality* of sensation means the quality of the sensation felt—that is, whether the sensation is one of pain, heat, cold, touch, position, or so forth. The part of the brain most important for determining the modality of sensation is the thalamus. When sensory impulses reach this part of the brain, those having to do with pain terminate in a different area from those having to do with touch, heat, and cold, and each of these terminates in other respective areas. The area of termination determines which modality of sensation will be felt.

THE AFFECTIVE NATURE OF SENSATIONS. In addition to determining the modality of sensation, the thalamus also determines whether or not the sensation is pleasant or unpleasant. This is called the *affective* nature of a sensation. When sensory signals reach the thalamus, parts of the signals are relayed into the *medial nuclei* located in the very middle of the thalamus. It is here that the affective characteristics are detected. For instance, pain is an unpleasant sensation while warmth is usually pleasant. On the other hand, extreme heat and cold are unpleasant. If the sensations are pleasant, the person becomes quite tranquil; if unpleasant, he usually becomes excited and takes steps to remedy the situation.

CONSCIOUSNESS IN THE THALAMUS. The idea that the thalamus contributes to consciousness is contrary to beliefs many scientists previously held about the function of the brain, for it was long believed that only the cerebral cortex operates at a conscious level. Yet the areas of the cerebral cortex having to do with somesthetic sensation can be extirpated, and the person will still be capable of detecting sensory modalities and their affective qualities. On the other hand, damage to the sensory areas of the thalamus destroys these abilities. Therefore, it must be concluded that at least the sensory part of the thalamus operates at a conscious level, and because of its very important abilities it is, considering its size, perhaps the most necessary part of the entire brain.

Pain

The stimulus that causes pain. Perhaps the most important of all sensations is that of *pain*, for this sensation helps us to protect the body from damage.

Pain receptors are stimulated when tissues of the body are *being* damaged. For instance, one feels pain while the skin is *being* cut, but shortly after the cut has been made, the pain generally is gone. Indeed, thousands of soldiers on the battlefield in World War II who had been mortally wounded were asked whether or not they felt pain. In most instances, no pain was actually felt except for a few minutes after the damage had been inflicted.

Different types of damaging stimuli that can cause pain are *trauma* to the tissues, *ischemia* of the tissues (lack of blood flow), intense *heat* to the tissues, intense *cold* (especially freezing of the tissues), or *chemical irritation* of the tissues. It is believed that as tissues are damaged they release some substance from the cells that stimulates the pain nerve endings. *Histamine* has been suggested as this substance, though this has not been proved.

PERCEPTION OF PAIN. It is frequently said that one person perceives pain more intensely than others. However, experiments in which graded intensities of tissue damage were caused in a large number of different persons showed that all normal persons perceive pain at almost precisely the same degree of tissue damage. For instance, when heat is used to cause tissue damage, almost all persons feel pain between the temperatures of 44 and 46° Centigrade, which is a very, very narrow range.

REACTIVITY TO PAIN. Even though almost all people perceive pain at the same intensity of tissue damage, not all people react alike to the same pain, for some react violently to only slight pain while others can withstand tremendous pain before reacting at all. This is determined not by differences in sensitivity of the pain receptors themselves but, instead, by differences in psychic makeup of the individuals. Therefore, when a person is said to be extremely "sensitive" to pain, it is meant that he *reacts* to pain far more than do other persons and not that he perceives far more pain.

Mechanism for Localizing Sensations on Specific Surface Areas of the Body

Among the most important information that the brain must determine about each somesthetic sensation is its point of origin on the body. Once

the location has been determined, it is possible to do something about the condition giving rise to the sensation. To provide this localization ability, the nerve fibers are *spatially oriented* in the nerve trunks, in the spinal cord, in the hindbrain, and in the cerebral cortex. This may be explained as follows.

Spatial orientation of nerve fibers. The sensory nerve fibers arising in the leg are separated in the spinal cord and brain from the fibers arising in the arm. The fibers arising from adjacent toes are separated from each other, and even the fibers arising from two areas of skin only one centimeter apart are kept separate from each other. In the thalamus, fibers from each area of the body end at a specific point. Therefore, the thalamus is capable of determining, though only roughly, which part of the body is being stimulated. The thalamus is not organized satisfactorily to localize sensations to very discrete areas of the body; instead, this function is performed in the somesthetic area of the cerebral cortex.

The somesthetic cortex. Figure 225 illustrates the *somesthetic cortex*, which is located immediately posterior to the central sulcus of the brain and extends from the longitudinal fissure into the fissure of Sylvius. This figure also shows the areas of the somesthetic cortex that receive sensations from each of the different parts of the body. Sensations from the foot, for instance, excite the somesthetic cortex where it dips into the longitudinal fissure. The leg area is approximately at the point where the somesthetic cortex comes out of the longitudinal fissure; then comes the thigh area, the abdomen, thorax, shoulder, arm, hand, fingers, thumb, neck, tongue, palate, and larynx. Thus, the entire body is spatially represented in the somesthetic cortex, each discrete point in this portion of the brain corresponding to a discrete area only a few millimeters in size on the body surface.

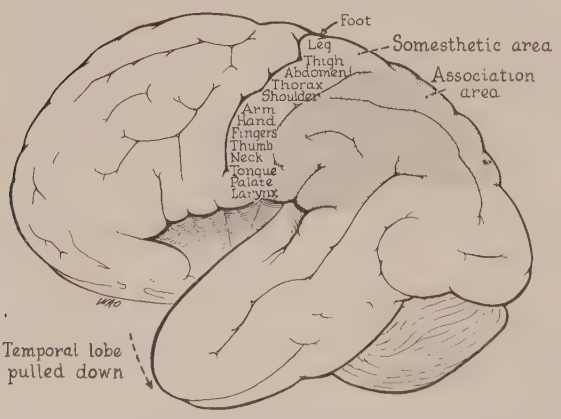


Figure 225. The somesthetic cortex and the somesthetic association area, showing also localization of sensations in the somesthetic cortex.

The function of the somesthetic cortex is to localize very exactly the points in the body from which the sensations originate. Though the thalamus is capable of localizing sensations to very general areas, such as to one arm, to a leg, or to the body, it is not capable of localizing sensations to minute places on the body. Instead, the thalamus relays the necessary signals into the somesthetic cortex where a much better spatial representation is available, and there the job of discrete localization is performed.

Not all types of sensations are localized equally well. The sensations of aching pain, crude touch, warmth, and cold are localized to general areas of the body rather than to discrete areas. It seems that the thalamus performs most of the localization function for these modalities of sensations. On the other hand, the sensations of light touch, pressure, and position are very discretely localized by the somesthetic cortex. These are the sensations normally detected by the special end-organs, for instance, Meissner's corpuscles for light touch and joint receptors for position sensation.

The somesthetic cortex is not capable of discriminating modalities of sensation. If one stimulates different portions of this cortex with an electrical stimulator, only a prickling sensation will be felt in the respective area of the body represented by the stimulated point. Only rarely will heat, cold, pain, or touch be felt. Therefore, the function of the somesthetic cortex is almost entirely to determine from which part of the body a sensation originates.

The dermatomes. Thirty large nerves called *spinal nerves* leave each side of the spinal cord and then spread to all areas of the body. Eight of these nerves originate from the neck portion of the cord, twelve from the chest portion, five from the lumbar portion (small of the back), and five from the sacral portion. Figure 226 illustrates the surface areas of the body subserved by each of the spinal nerves, and each of these areas is called a *dermatome*. It will be noted that many of the nerves originating in the neck innervate dermatomes all the way down to the arm and hand, and that lumbar and sacral dermatomes extend downward into the leg and foot.

When sensation has been lost in some part of the body, the dermatomal chart of Figure 226 can often be used to determine what portion of the nervous system has become non-functional. For instance, if sensation in the anterior portion of the

thigh is gone, one can reason by observing the dermatomal chart that one of the lumbar nerves or the lumbar portion of the cord itself has been damaged.

Discrimination of Intensity of Sensations—The Weber-Fechner Law

A person can detect the weight of a flea on the tip of his finger or he can detect the weight of an elephant stepping on the same finger. The difference in weight of these two animals is approximately nine billion times, and yet, from the sensations perceived, a person can estimate the heftiness of the two objects. The reason why one can discriminate such wide differences in intensities is that the number of impulses transmitted by a sensory nerve is roughly proportional to the logarithm of the intensity of sensation rather than to the actual intensity. In the case of the flea and the elephant the difference between the logarithms of their weights is only about ten times even though the difference between the actual weights is nine billion times. This logarithmic detection of sensation applies not only to somesthetic sensations, but also to the sensations of seeing, hearing, taste, and smell; it is called the *Weber-Fechner law.*

Another example of the Weber-Fechner law is the following: If one is holding an object in his hand that weighs 1 ounce and he exchanges this object for another that weighs 1.1 ounce, he will be able barely to discriminate the difference in weight of the two objects. The actual difference in weight is 0.1 ounce. Then he picks up an object that weighs 10 pounds, and exchanges this object for another that weighs 10 pounds and 0.1 ounce. He will be completely unable to discriminate the difference in weight. Instead, the second object must now weigh 11 pounds for him to tell the difference. The increase in weight this time must be 1 pound instead of 0.1 ounce. In each instance, the increase is 10 per cent of the original weight. Therefore, another way of expressing the Weber-Fechner law is: discrimination of intensity of the different types of sensation is on relative basis rather than on an absolute basis.

THE VISCERAL SENSATIONS

The visceral sensations are those that arise from the internal structures of the body, such as the organs of the abdomen and chest, but also include sensations from inside the head, from the muscles, the bones, and other deep structures.

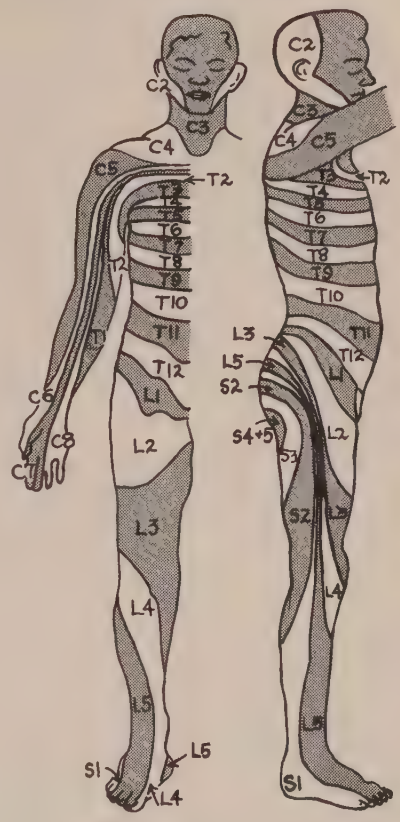


Figure 226. Sensory areas (dermatomes) on the surface of the body subserved by the respective spinal nerves. (Modified from Grinker and Bucy: *Neurology.* Charles C Thomas.)

The usual modalities of visceral sensation are pain, burning sensations, and pressure (which is often manifested as a sensation of fullness). Because these modalities are also exhibited by the exteroceptive sensations, the visceral sensations are sometimes considered to be part of the exteroceptive system. Indeed, the pathways of these two types of sensation are closely related, as will be noted below.

Pathways for visceral sensations. Ordinarily one is not at all conscious of his internal organs, but an inflamed organ can transmit pain sensations to the central nervous system through two separate pathways, the *parietal pathway* and the *visceral pathway,* both of which are illustrated for sensations from the appendix in Figure 227. The parietal pathway is the same as the pathway for transmission of exteroceptive sensations. In the case of the appendix, the inflamed appendix irritates the peritoneum overlying the appendix, and pain impulses are transmitted from the peritoneum through the same nerve trunks that carry exteroceptive sensations from the outside of the abdominal wall at this point.

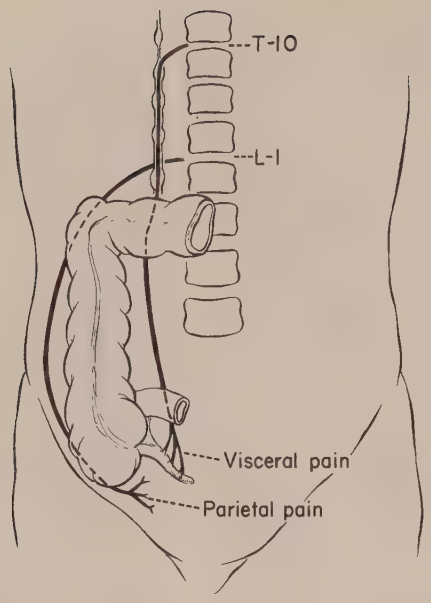


Figure 227. Parietal and visceral pathways for transmission of pain from the appendix.

The visceral pathway utilizes sensory nerve fibers in the autonomic nervous system. Figure 227 illustrates sensory fibers leaving the appendix to enter the sympathetic chain. After traveling upward a few segments the fibers leave the sympathetic chain to enter the spinal cord at approximately the T-10 segment in the lower thoracic region. Therefore, the visceral sensations from the appendix enter the spinal cord at an entirely different point from the parietal sensations.

Referred pain. Pain from an internal organ is often felt on a surface area of the body rather than being localized in the organ itself. This is called *referred pain.* It may be referred to the surface immediately above the organ, or often to areas considerable distance away. The mechanism of referred pain is probably that illustrated in Figure 228, which shows two visceral nerve fibers as well as two additional nerve fibers from the skin entering the spinal cord. Both of these types of fibers synapse with the same two neurons in the spinal cord. Therefore, stimulating either the visceral fibers or the skin fibers will send impulses up the spinal cord along the same pathways. Because the person has never had reason to know the location of his internal organs but is very familiar with the location of the different skin areas, the impulses from the visceral fibers are usually interpreted as coming from the skin.

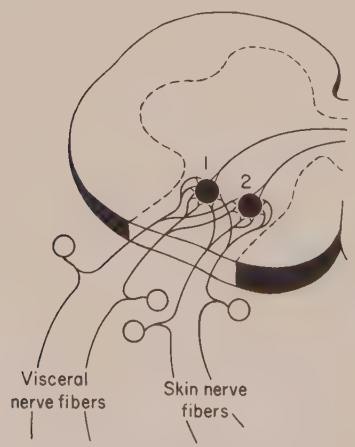


Figure 228. The neurogenic mechanism responsible for referred pain.

When a visceral sensation is transmitted through a parietal pathway, the referred pain is usually felt on the surface of the body directly over the respective internal organ, but when the sensation is conducted through a visceral pathway, it is usually referred to a skin area quite remote from the organ. The reason for this is that fibers of the visceral pathway usually travel a long distance in the sympathetic chain before entering the spinal cord. Figure 229 illustrates the surface areas to which pain is referred from many of the different organs. For instance, though the appendix lies far to the right and in the lower abdomen, its visceral sensations are referred to an area around the umbilicus. Referred pain from the kidney and ureter occurs near the midline on the anterior abdominal wall, though these organs actually lie in the posterior part of the abdomen.

Visceral pain from the heart is among the most important of the referred sensations. Figure 229 shows the areas to which cardiac pain is often referred, including the upper thorax, the shoulder, and the medial side of the left arm, particularly along the radial artery.

The surface area to which visceral pain is referred usually corresponds to that portion of the body from which the organ originated during embryonic development. For example, the heart originates in the neck of the embryo and so does the arm; therefore, heart pain is frequently referred to the arm. The appendix, as another example, originates from the portion of the primitive gut that develops near the umbilicus, which explains the reference of appendiceal pain to the umbilical region. The other areas of referred pain illustrated in Figure 229 also correspond with the areas of origin of the different organs.

The stimulus for visceral pain. Cutting through the gut, the heart, the liver, the muscle, or other internal organs with a sharp knife causes almost no pain. Instead, pain from these areas is produced much more easily by (1) stopping the

blood flow to the area, (2) application of an irritant chemical over wide areas, (3) stretching the tissues, or (4) spasm of the muscle in the organ.

One of the most important stimuli for visceral sensation is *ischemia,* which means lack of blood flow. The pain of a heart attack, as an example, is caused by poor flow through the coronaries to the heart muscle, and one of the most intense pains ever felt from the abdominal viscera is that caused by blockage of one of the major arteries supplying the gut. Even the skeletal muscles become extremely painful when their blood supply is diminished for a prolonged period of time. The reason for the pain caused by ischemia has never been determined, though it is believed that the lack of blood flow allows metabolic end-products to build up in the tissues, and that these in turn produce irritant effects on the pain nerve endings.

Stretching the tissues possibly causes pain by producing ischemia, for overstretching the tissues occludes the blood vessels. Likewise, spasm of the muscle in an organ can compress the blood vessels; spasm also increases the rate of metabolism so that far more than usual quantities of metabolic end-products are dumped into the tissues. These factors, therefore, could explain why spasm of the gut often causes very intense abdominal pain.

Qualities of visceral pain sensations. Visceral pain sensations are usually of an aching, burning, or cramping nature and spread over a wide area. The reason why a person generally feels a widespread pain rather than a sharp pain is probably that the viscera are not supplied with a great number of pain nerve fibers. Therefore, before the pain can be felt, a large area usually must be stimulated at once. Mild stimulation of the fibers is generally felt as an aching pain, but as the stimulation becomes more intense the pain develops a burning quality. Therefore, the difference between stimulation producing an aching sensation and that producing a burning sensation is probably not a qualitative difference, but instead only a difference in intensity.

Cramping pain appears as a slow rhythmic cycle, becoming very intense for a while, then disappearing, and returning again and again. It can be either aching or burning, but more frequently is the former. The cause of cramps is usually intermittent spasm of the smooth muscle in an organ, spasm resulting from overactivity of its normal rhythmic contractions.

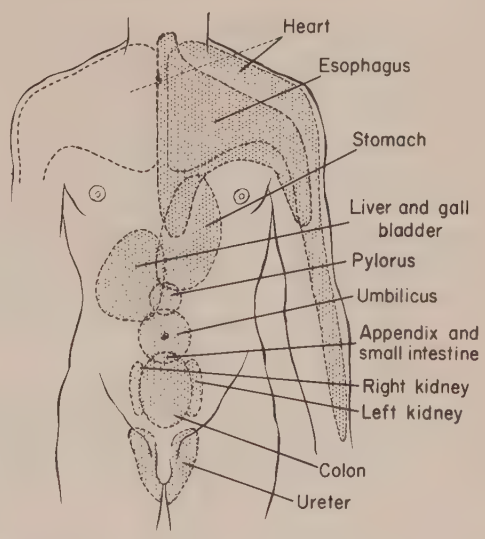


Figure 229. Surface areas of the body to which pain from different organs is referred.

Headache

Headache is another type of referred pain, and it is usually caused by irritation or damage occurring in the tissues inside the head. Figure 230 illustrates the surface areas of the head to which pain from the deep structures is referred. Pain fibers from inside the eye and from the nasal sinuses are transmitted through the first and second divisions of the fifth nerve. This nerve also supplies the skin areas over the lower forehead and around the eye and nose. Therefore, sinus infections or irritation of the eyes caused by overuse or by intense light can result in dull aching pain referred diffusely over the frontal and orbital areas of the head.

Irritation occurring inside the skull but above the level of the ears—that is, above the tentorium —causes headache referred to the frontal and temporal surfaces of the head. The reason for this is that the third division of the fifth cranial nerve

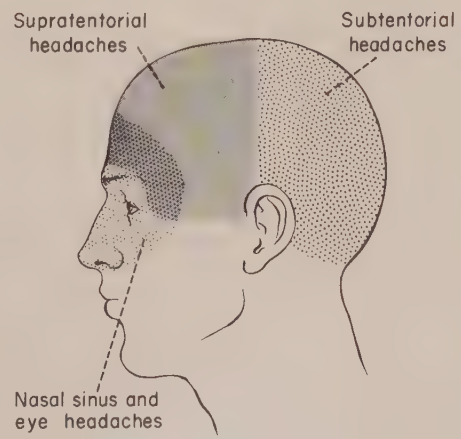


Figure 230. Surface areas of the head to which different types of headache are referred.

supplies both the supratentorial areas inside the skull and the surface areas of the skull in the temporal and frontal regions.

Irritative effects beneath the tentorium, in the pocket of the skull beneath the level of the ears, give rise to headache localized over the occipital part of the skull, also shown in Figure 230. Both the subtentorial areas inside the cranial vault and the occipital surface areas of the skull are supplied by the upper cervical spinal nerves, which explains the reference of subtentorial pain to the occipital regions.

The intracranial stimulus for headache. Headache originating inside the skull is usually caused by irritation of the *meninges,* which are the membranes surrounding the brain. Some causes of very severe headache of this type are: (1) meningitis, an infection of the meninges, which causes very intense headache; (2) removal of fluid from the spaces around the brain, allowing the brain to rub freely against the meninges and to irritate them, this too causing very intense headache; and (3) an alcoholic binge, which probably causes headache afterward by irritation of the meninges, for the meninges almost certainly become reddened and inflamed in the same manner as the whites of the eyes the day after an alcoholic bout.

Headaches usually do not result from pain originating in the substance of the brain itself. In fact, a patient who is having a brain operation under local anesthetic—that is, without being asleep—can feel no pain when the surgeon cuts through the brain tissue. Yet when he cuts the meninges, or especially when he cuts one of the major blood vessels supplying the meninges, intense headache is experienced.

THE INTERPRETATION OF SENSATIONS BY THE BRAIN

Interpretation of Somesthetic Sensations—The Somesthetic Association Area

Figures 223, 224, 225, and 231 all illustrate an area of the cerebral cortex, immediately posterior to the somesthetic cortex, called the *somesthetic association area.* Impulses pass into this area from several sources: (1) from the somesthetic cortex, (2) directly from the thalamus, and (3) from various other parts of the cerebral cortex.

The function of the somesthetic association area is to interpret the "meaning" of the somesthetic sensations arriving in the brain. It is here that the brain determines the shapes of objects, their weights, the textures of their surfaces, and their positions in relation to the body. This area also keeps track of the exact position of each part of the body at all times.

Many of the memories of past sensory experiences are stored in the somesthetic association area, and when new sensations similar to the old ones arrive in the brain the parallel nature of the two sensations is immediately discerned. It is in this manner that one associates a new sensation with previous ones. As more and more sensory experiences accumulate, new sensory experiences can be interpreted on the basis of what is remembered from the past.

Cortical Localization of Other Sensations

The sensations of vision, hearing, taste, and smell, like the somesthetic sensations, are each transmitted from the receptor organs to small circumscribed areas in the cerebral cortex, each of the areas having a different location. These areas are called the *primary cortical areas* for sensations, and from each primary area the impulses go to an *association area.* The primary area for somesthetic sensation, the somesthetic cortex, and the somesthetic association area were discussed above.

Cortical areas for vision. The *primary visual cortex* is located in the *calcarine fissure area* in the posterior part of the brain on the medial side of each hemisphere. A small portion of the primary visual cortex extends over the occipital pole, as shown in Figure 231, though most of it is hidden inside of the longitudinal fissure. The primary visual area interprets only a few of the meanings of visual sensations. It is mainly a relay station for transmitting the visual image.

Nerve fibers pass in all directions from the primary visual cortex into adjacent areas called the *visual association area,* which also is shown in Figure 231. This area interprets the meaning of the visual signals. It interprets the shapes and the colors of the different objects and then identifies the objects. Finally, the association area interprets the overall meaning of the scene before the eyes.

Interpretation of the written language is one of the most important functions of the visual association area. To accomplish this feat, this area must first discern from the light and dark spots of the letters themselves, then from the combination of letters the words, and from the sequence of words the meaning of the thoughts.

Cortical areas for hearing. Auditory sensations are transmitted from the ears to a small area called the *primary auditory cortex* in the upper part of the temporal lobe. From this area impulses

pass into the surrounding *auditory association area,* shown in Figure 231. The primary auditory cortex receives the sound impulses from the ears, and the auditory association area determines their meanings. One part of the association area determines whether the sound is noise, music, or speech; then other parts determine the thoughts conveyed by the sound. To interpret the meaning of speech, the auditory association area first combines the various syllables into words, then words into phrases, phrases into sentences, and finally sentences into thoughts.

Cortical areas for smell and taste. The *primary cortex for smell* is located on the bottom surface of the brain, possibly in the region of the *uncus,* though it has never been located exactly, and the *primary cortex for taste* is located at the bottommost end of the primary somesthetic area, deep in the fissure of Sylvius. From these primary areas impulses pass into surrounding *smell and taste association areas,* and there the sensations are interpreted in the same manner as somesthetic, visual, and auditory sensations are interpreted by their respective association areas.

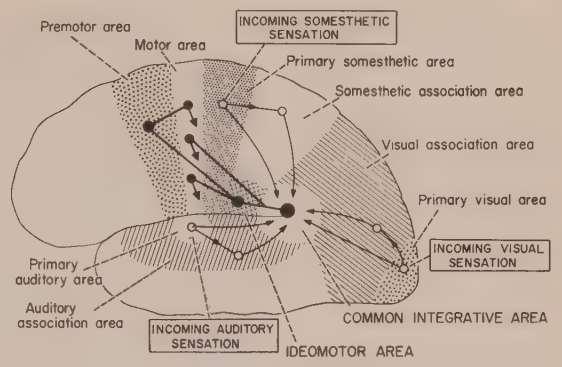


Figure 231. Integration of sensory signals from several different sources into a common thought by the common integrative area of the brain, showing also the primary and association areas for vision, for auditory sensations, and for somesthetic sensations.

The Common Integrative Area of the Brain—the Gnostic Area

Figure 231 illustrates the passage of signals from the somesthetic, visual, and auditory association areas into a *common integrative area,* located in the angular gyrus of the brain midway between these three respective association areas. Signals are also transmitted into this area from the taste and smell association areas. It is here that all the different types of sensations are integrated to determine a common meaning, which is the reason for its name, *common integrative area.* This region of the brain is also called the *gnostic area,* which means the "knowing area." If a person were in a jungle and heard a noise in the brush, saw the leaves moving, and smelled the scent of an animal, he might not be able to tell from any one of these sensations exactly what was happening, but from all of them together he could quite readily assess his danger. It is in the common integrative area that all the thoughts from the different sensory areas are correlated and weighed against each other for deeper conclusions than can be attained by any one of the association areas alone.

Most of the sensory information arriving in the brain finally is funneled through the common integrative area. For this reason, any damage to this area is likely to leave the person mentally inept; even though the different association areas might still be able to interpret their respective sensations, this information is almost valueless to the brain unless its final meaning can be interpreted. Some signals can be transmitted directly from the association areas to other portions of the brain without going through the common integrative region, but these are so few that the person who loses his entire common integrative area generally becomes an imbecile. This is occasionally the unfortunate result of a brain tumor or a *stroke.* (A stroke is caused by sudden loss of blood supply to an area of the brain because of hemorrhage or thrombosis of a blood vessel.)

Dominance of one side of the brain. The common integrative area is located in the angular gyrus of the left cerebral hemisphere in at least nine tenths of all people. The angular gyrus of the opposite hemisphere is usually almost totally nonfunctional, though in one tenth of the people it is the common integrative area and the left angular gyrus is non-functional. At birth both angular gyruses probably have almost equal functional abilities, but the right angular gyrus region usually becomes almost totally suppressed as the brain develops, while the left angular gyrus becomes the most important portion of the entire cerebral cortex. Occasionally, though, the relationship between the two is reversed. This phenomenon of *suppression* occurs very commonly in different parts of the brain when signals from two different sources interfere with each other. For instance, when a person is cross-eyed, so that the visual signals from the two eyes do not correspond satisfactorily, the brain automatically suppresses the sensations from one of the eyes. As a result, this eye gradually becomes "functionally" blind until almost all vision is lost, while the person develops excellent vision in the other eye.

Control of motor functions by the common integrative area. Once the common integrative area has integrated all incoming sensations into a common thought, signals are then sent into other portions of the brain to cause appropriate responses. Most of the signals pass to the *ideomotor area,* an area immediately anterior to the common integrative area. Here the course of action to be taken is determined in response to the incoming signal. If the integrated thought indicates that muscular activity is needed, the ideomotor area sends impulses into the premotor and motor portions of the brain to cause muscular contractions. If muscular activity is not called for, the ideomotor area simply initiates thoughts throughout different portions of the brain. All these reactions of the ideomotor area, premotor area, motor area, and thought regions of the brain are discussed at greater length in the following chapter.

Function of the thalamus in association with the cortex. Earlier in the chapter it was pointed out that the modality of sensation—that is, the type of sensation—is interpreted principally by the thalamus. Therefore, the thalamus must be involved as much as the cortical regions of the brain whenever sensory signals enter the brain. Indeed, all cortical functions are probably dependent upon simultaneous thalamic function, for each cortical area is connected directly with a corresponding area in the thalamus, to and from which nerve impulses maintain constant communication. When an area of the cortex becomes activated, the corresponding area in the thalamus also becomes activated. Therefore, whenever we talk about interpretation of sensations in the cortex and about transmission of sensory signals from one cortical region to another, we must realize that corresponding transmission of signals occurs at the same time in the thalamus.

It is also believed that cortical function is actually *controlled* by the thalamus, that the awake state in the cortex is caused by impulses transmitted through the thalamus and that signals from the thalamus activate specific portions of the cortex thereby calling one's attention to the type of information stored in the respective portion of the cortex. To express this another way, the thalamus seems to be the main *control center* for eliciting cortical activity, while it is in the cortex that most memories are stored and where incoming sensory signals are compared with the memories so that their meanings can be interpreted.

REFERENCES

Deutsch, J. A.: Higher nervous function: the physiological bases of memory. *Ann. Rev. Physiol.* 24:259, 1962.

Friedman, A. P., editor, and Merritt, H. H.: *Headache: Diagnosis & Treatment.* Philadelphia, F. A. Davis Company, 1959.

Hardy, J. D., Wolff, H. G., and Goodell, H.: *Pain Sensations and Reactions.* Baltimore, The Williams & Wilkins Company, 1952.

Horridge, G. A.: Comparative physiology: integrative action of the nervous system. *Ann. Rev. Physiol.* 25:523, 1963.

Perl, E. R.: Somatosensory mechanisms. *Ann. Rev. Physiol.* 25:459, 1963.

Pieron, H.: *The Sensations: Their Functions, Processes, and Mechanisms.* London, Frederick Muller, Ltd., 1952.

Rose, J. E., and Mountcastle, V. B.: Touch and kinesthesis. *Handbook of Physiology* 1:387, 1959.

Sweet, W. H.: Pain. *Handbook of Physiology* 1:459, 1959.

Weddell, G., and Miller, S.: Cutaneous sensibility. *Ann. Rev. Physiol.* 24:199, 1962.

Zotterman, Y.: Thermal sensations. *Handbook of Physiology* 1:431, 1959.

THE THOUGHT PROCESSES AND CONTROL OF MOTOR ACTIVITIES

24

THE NATURE OF THOUGHTS

A thought may be defined as a pattern of neuronal impulses traveling in certain pathways in the brain. If the impulses travel along one pathway the person has one thought, but if the pathway is different, the thought is different. Figure 232 shows a postulated type of thought pattern. In this figure an incoming sensory impulse arrives in the thalamus and causes secondary impulses to be transmitted, first, to neuron station number 5, then to station number 1, to number 2, to number 3, to number 4, back to number 5, and around and around the circuit again and again. As long as the impulses continue to reverberate around this circuit the person continues the same thought. This schema, however, is extremely simplified, for many collateral nerve branches spread from each of the neurons to wide-reaching areas of the brain, all of which undoubtedly act almost simultaneously to produce the thought.

The tendency for thoughts to change. After the circuit in Figure 232 has reverberated for a short period, it is reasonable to expect that at least one of the cells in the circuit will become fatigued. As a result, the reverberation will stop and the thought will cease. Then a second thought

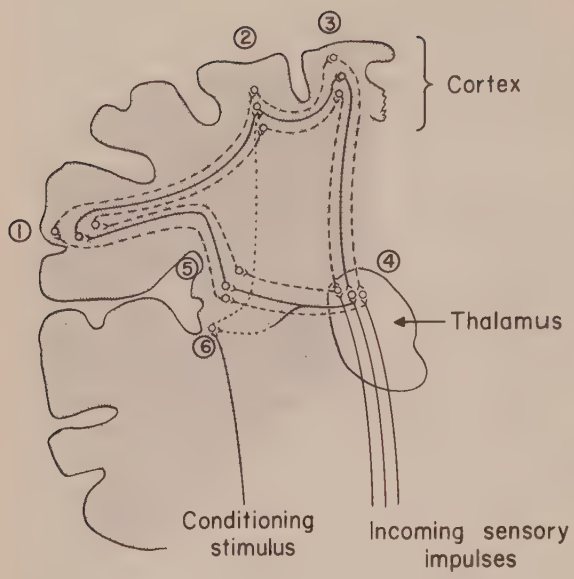


Figure 232. A theoretical mechanism for thoughts in the brain.

pattern might be initiated by other incoming sensory impulses.

Another method by which thoughts can change from moment to moment is for the reverberating thought pattern to assume a slightly different pathway from the original, with the reverberation never ceasing. This might result from a conditioning stimulus. For instance, in Figure 232 a conditioning stimulus might facilitate neuron number 6 so that collateral impulses arriving from the reverberating thought pattern would then be able to stimulate this neuron. As a result, a new reverberating pathway would now be initiated through neurons 2, 3, 4, and 6, by-passing neurons 5 and 1. This would be a slightly different pattern from the original and would give a closely allied though different thought. This procedure for changing thoughts is probably very common, because most thoughts do seem to drift to another similar thought rather than to change completely all at once.

Intensity of thoughts. The intensity of a thought is probably determined by the number of parallel neurons involved in the thought process. For instance, in the thought pattern of Figure 232, if the reverberating pathway utilizes only the neurons connected by the solid line in the figure, only these few neurons will be involved in the pattern, and the intensity of the thought will be weak. If adjacent neurons are stimulated simultaneously and these join in the reverberating pathway so that many parallel fibers are excited simultaneously, then the intensity of the thought will be greatly increased.

It must be emphasized again that these ideas about the nature of thoughts, including the mechanisms for intensity of thoughts, changing thoughts, and so forth, are hypothetical, though electrical studies from the brain during thought processes do indicate that mechanisms at least similar to these are responsible for the different thoughts. Also, recent experiments have shown that every time the cortex is involved in a thought process, certain centers deeper in the brain are equally as much involved. In the previous chapter it was pointed out that every area in the cortex is directly connected to a corresponding area in the thalamus and that signals constantly reverberate back and forth between these two areas. In addition, regions of the mesencephalon, hypothalamus, and basal ganglia are connected with the cortex and involved in certain aspects of thoughts.

Memories

Memories are thoughts that recur seconds, minutes, or even years after the same thought had first occurred. When a person sees a scene, for example, his eyes transmit a certain pattern of impulses into the primary visual cortex, then into the visual association areas, into the common integrative region of the brain, and to many other areas of the brain. This pattern of impulses creates a thought that describes the scene before his eyes. The person can then close his eyes and continue to project the scene in his mind. This is a memory. The ability to project in one's mind scenes, sounds, abstract facts, and other patterns of mental activity, without the necessity of immediate stimulation from sensory impulses, is memory.

The memory process. Recent research studies indicate that the basis of the memory process is the ability of the synapses to become *permanently facilitated* when impulses pass through them. As described in the preceding paragraph, when a person sees a scene and a particular thought process is created, the passage of impulses from the eyes through the visual cortical areas and then through the other areas of the brain lowers the threshold for stimulation of these same pathways by signals attempting to pass along them at a later date. That is, passage of an impulse across a synapse causes some effect at the synapse that allows second, third, fourth, and more impulses to pass through the same synapse with progressively greater ease.

The more times a particular pathway is used, the more facilitated it becomes. After the same thought process has been evoked many times by

the same incoming sensory impulses, the pathway finally becomes so facilitated that incoming sensory impulses are no longer needed to elicit the thought. Instead, spontaneous signals arising in the brain can feed into the facilitated pathway, setting off the thought. As a result, exactly the same scene that the person had previously seen with his eyes is projected in his mind. The same is true for thoughts initiated by hearing sounds, feeling objects, smelling odors, tasting food, and thinking abstract thoughts.

Fabrication. Fabrication means the development in one's mind of thoughts differing from those originally initiated by incoming signals. Usually this occurs by mixing memories together. A portion of one memory process is added to a portion of another, and these then are added to still other partial memories until one comes out with a completely new thought that had never been directly stimulated by the incoming signals. Yet, these thought processes are an indirect result of original stimulation of the sensory receptors and of the storage of memories.

Another method of fabrication occurs especially in old age. The elderly person often tells wild tales and honestly believes that he is recounting true experiences of the past. This probably results from aging of the memory pathways. As the brain becomes older, neurons are often destroyed, and when this happens the thought pathway finds a detour through other neurons, thus changing the thought. Also, spurious conditioning stimuli from any source can probably condition neurons immediately adjacent to the thought pathway and thereby allow the thought pattern to take a slightly new route without the person being aware of the changing nature of the thought. Yet, as the thought is recalled again and again, it becomes more and more deeply ingrained in the person's memory even though it is mutating gradually from the original, sometimes becoming nothing more than a "tall tale." This probably explains why many older people have very vivid "memories" of their childhood days though the truth might actually be quite different from the memories.

Knowledge

Knowledge is the accumulation of memories in the brain. There are many different types of knowledge. For instance, in the visual association areas many visual memories accumulate, in the auditory association areas many memories of different combinations of sounds accumulate, and the same is true for tactile memories in the somesthetic association areas, olfactory memories in the olfactory association areas, and gustatory memories in the taste association areas. In addition to these, memories of abstract thoughts are stored in the common integrative region of the brain, in the temporal lobes, and in the prefrontal lobes. These memories can be recalled at any time by sending impulses into the appropriate area of the brain, thereby setting off the reverberatory process of the desired memory.

The total amount of knowledge that a person accumulates in his brain is determined by his rate of learning and his rate of forgetting. In the earlier years of life a person learns very rapidly and forgets little. In the later years of life he often forgets as rapidly as he learns and, in many instances, even more rapidly than he learns. Psychological tests for all types of knowledge indicate that the average person reaches a maximum accumulation of all types of knowledge at an age of approximately 39 years. This does not mean that his knowledge for any specific subject will be greatest at this age, for if one begins studying a subject at the age of 30, he may attain the age of 50 or 60 before he reaches his maximum accumulated knowledge in that particular field. But at the same time that he is learning this new subject, he will be forgetting many of the things that he had learned in childhood. For instance, the knowledge in the brain that controls the muscular functions reaches peak development in the late teens and early twenties. Much of this knowledge will be progressively lost in the thirties and forties while the person is learning new things.

THOUGHT PATTERNS IN THE BRAIN FOR CONTROLLING MOTOR ACTIVITIES

Large portions of the brain are concerned specifically with controlling the muscles of the body. These brain areas are shown diagramatically in Figure 233. The so-called *motor area* lies immediately anterior to the central sulcus—that is, directly in front of the somesthetic cortex. This area controls the actions of the discrete muscles throughout the body. Immediately anterior to the motor area is the *premotor area,* which stimulates the motor area to produce coordinate movements of many muscles at once.

Located anterior to the common integrative region of the brain and posterior to the primary somesthetic cortex is the *ideomotor area.* This area initiates complicated patterns of muscular activity, and it carries these out by controlling both the motor and premotor areas.

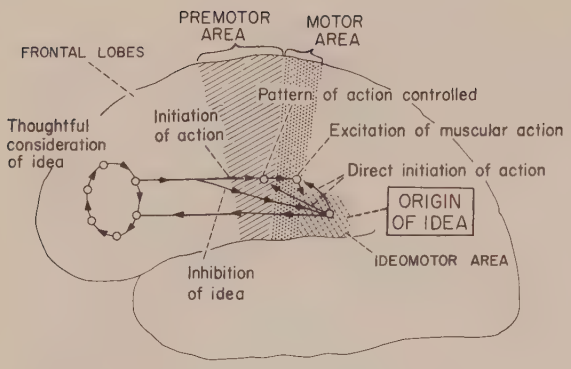


Figure 233. Pattern of impulses in the brain for originating and controlling motor functions.

Finally, also of importance in the control of the muscles is the *prefrontal area* of the brain which is comprised of both anterior poles of the brain called the *prefrontal lobes.* This area enters into abstract thinking as well as into the immediate control of muscular activity. It is the area in which considered judgment occurs before acting. It operates in conjunction with the ideomotor and common integrative areas to help determine the sequence of muscular functions as well as the sequence of other thought processes of the brain.

Function of the Ideomotor Area

In the preceding chapter it was pointed out that all of the sensory signals from the different sensory receptors are finally integrated into coordinate thoughts in the *common integrative* or *gnostic* region of the brain, located in the angular gyrus where the parietal, temporal, and occipital lobes of the brain all come together. Immediately anterior to this region, in the supramarginal gyrus of the parietal lobe, is a small area called the *ideomotor area.* This is very closely associated with the common integrative region, for its function is to decide on a course of action in response to the coordinated information from the common integrative region. For instance, if an adversary strikes one in the face with his fist, the vision of the fist coming toward the face, the pain of the fist striking, and the sound of the nose crunching all immediately give the person a vivid picture in the common integrative region of his brain of an enemy that needs to be repelled. The ideomotor area determines that the best procedure is to strike the adversary back as rapidly and as hard as possible. As a result, it sends a sequence of impulses to the motor regions of the brain to cause the fist to be doubled, the arm to be lifted, to be drawn back, and then the fist to be moved forward until it strikes. All of these activities are very rapid and initiated almost automatically on the part of the ideomotor area.

The ideomotor area ordinarily does not delay long in making decisions. It considers very few factors other than the immediate incoming sensory information. It usually forgets about the memories of previous battles with adversaries too big to fight, and it reacts rapidly and quickly without considered judgment.

It is the ideomotor area that allows a person to answer questions quickly and snappily. For instance, if someone asks the product of 2 times 2, the person immediately says 4, or, if he is asked to add 10 and 10, he says 20. But when any complicated mathematical or other type of question is put to him, his ideomotor area is incapable of holding the thought for a long enough period of time to work out the solution. For instance, if he is asked to add in his head 157 plus 233, it would be impossible for him to perform this mental gymnastic without aid from other areas of the brain.

Dominance of the ideomotor area. The ideomotor area, like the common integrative area, is almost always located on the left side of the brain, and the corresponding area of the opposite side of the brain is suppressed. This presumably prevents confusion between two separate areas trying to control the sequence of motor activities at the same time.

Function of the Prefrontal Area

The prefrontal lobes operate in association with the ideomotor area to allow considered judgment before acting in response to incoming impulses. If the response of the person is decided entirely in the ideomotor area alone, only a rapid and inconsiderate reaction is likely to occur. But if the prefrontal area is operating simultaneously, the person will pause a moment and consider the possible consequences of his reaction before responding. To do this the prefrontal area has two specific control functions over the ideomotor area. These are, first, a suppressor function and, second, an excitatory function.

Suppression of the ideomotor area. The prefrontal area keeps the ideomotor area suppressed most of the time. As a result, regardless of the incoming sensory signals, the ideomotor area, though capable of responding rapidly and violently, is inhibited in its response until the prefrontal and other thought areas of the brain can have time to think over the type of response that

will be most advantageous. The common integrative region transmits its sensory information to the prefrontal lobes, to the temporal lobes, and to the memory circuits of the different association areas. In these areas, thought processes are initiated, and memories are immediately called forth of similar situations in the past. Then, these memories seem to funnel their information into the prefrontal lobes, where it is all coordinated into one master thought process.

Excitation of the ideomotor area. Once the prefrontal region has considered all the evidence both from the incoming sensory impulses and from previous experience, it then functions in unison with the other areas of the brain to determine the action most likely to lead to a satisfactory outcome. This constitutes the excitatory function of the prefrontal lobes. It is likely that the person will run when he is hit in the face by a large adversary but will strike back if the adversary shows all the evidences of being conquerable.

Other functions of the prefrontal area. The prefrontal lobes are also the loci of most of the abstract thoughts of the mind. It is principally the prefrontal area that is responsible for a person's ambition, his conscience, his planning of his future, and his worrying about details in order to achieve better efficiency in his activities. In other words, the prefrontal lobes seem to be continually concerned with the welfare of the individual, and they are always calling forth information from every part of the brain, sorting this information, attempting to find the secrets of better ways of living. It is also the prefrontal area that is capable of performing rigorous mental gymnastics such as solving difficult mathematical problems, and this area is the depository of the considered judgment necessary to a legal authority or a good medical diagnostician.

The difference between the function of the prefrontal lobes and other areas of the brain is their ability to store and utilize hundreds of small "bits" of information all at the same time. The storage of these bits is usually temporary, for only a few minutes at a time, but so long as the bits are present in the frontal lobes they can be combined together to form complex thoughts. For instance, an animal without prefrontal lobes that is exposed to a series of sequential experiences loses rapidly the information gained from moment to moment. He might forget the presence of a plate of food on the opposite side of the cage if he is not looking at it. Similarly, a human being without prefrontal lobes, on dismantling a watch, loses all concept of the positions of the parts, or on attempting to solve a complicated problem in logic he loses most of the information that is required for solving the problem before he is anywhere near the solution.

Effect of loss of the prefrontal lobes. Occasionally the prefrontal lobes are completely destroyed by disease or trauma, and sometimes they are destroyed purposefully because harmful patterns of thought develop that cannot be stopped in any other way. When the prefrontal areas are gone, the ideomotor and common integrative regions are left mainly in charge of the thought patterns and motor responses of the brain. The person without prefrontal lobes is likely to exhibit extreme reactions to sensory impulses, some of which are very happy in nature while others have the characteristics of extreme temper. In all instances the responses are rapid and are likely to lead to disastrous results. As long as the person is unprovoked, he has very much the same personality as a giddy teenager responding with feeling to everything but without much thought. When provoked he is likely to fall into a state of rage which cannot be easily quelled at the moment, but which will be totally forgotten a few moments later.

Function of the Motor and Premotor Areas of the Cortex

The motor cortex. The *motor cortex* is located in front of the central sulcus of the brain immediately anterior to the somesthetic cortex, as illustrated in Figure 234. This figure also shows several other areas that are associated with motor function, such as areas in the posterior part of the brain that help to choose the words for phonation and help to control eye fixation. Anterior to the motor

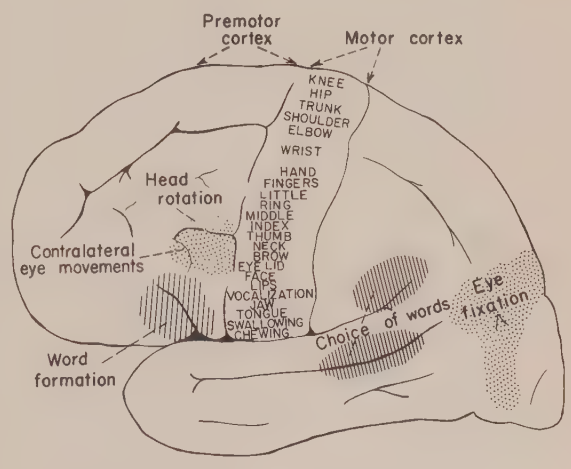


Figure 234. Representation of the different muscles of the body in the motor cortex, and location of other cortical areas responsible for certain types of motor movements.

cortex are areas that control automatic head rotation, contralateral eye movements, and the actual process of word formation. These areas are part of the *premotor cortex,* which is described below.

The primary function of the motor cortex is to control discrete movements of the different skeletal muscles in the body. Stimulation of a point area in the motor cortex causes a specific muscle on the opposite side of the body to contract. However, not all muscles are controlled to the same degree of precision by the motor cortex, for stimulation of a tiny area of the motor cortex may cause half of the trunk muscles on the opposite side of the body to contract all at the same time, while stimulation of another small area equally large may cause only one of the small thumb muscles to contract. Thus the "degree of representation" in the motor cortex is more than 100 times as great for the thumb as for the trunk muscles. Figure 235 illustrates schematically the degree of representation of the different muscles, showing that the hand and mouth muscles constitute about two thirds of the entire representation in the motor cortex. Note also that the very top of the motor cortex controls the toe, ankle, and foot muscles on the opposite side of the body, while the lower part of the motor cortex controls the face muscles on the opposite side.

The premotor cortex. The *premotor cortex,* illustrated in Figure 234, lies anterior to the motor cortex, between it and the prefrontal lobes. This area controls coordinate movements of several muscles at a time. It sends impulses either (a) to lower centers of the brain, which in turn transmit signals to the muscles or (b) to the motor cortex which in turn excites the individual muscles.



Figure 235. Degree of representation of the different body muscles in the motor cortex. (From Penfield and Rasmussen: *The Cerebral Cortex of Man.* The Macmillan Company.)

The premotor area seems to control "patterns" of movement. If one will stop to think for a moment, most of the motions he performs with his hands or other portions of his body fall into no more than a few hundred patterns of motion, but these are used over and over again in different sequences. Each time a particular pattern of movement is used, a neuronal circuit in the premotor cortex seems to become progressively more facilitated until eventually the movement can be performed rapidly and precisely without fumbling. This is theoretically the learning process for control of muscular activities. The reason for believing that the premotor area is such a locus for these circuits is that ablation of the premotor cortex causes loss of coordinate movements but does not cause loss of discrete movements of muscles. The greater the number of patterns of movements stored in the premotor area and the more precise they become, the more skilled the person becomes in the performance of physical tasks.

Overall control of motor activity. Most of what we know about the overall control of motor activity is based on studies in human beings who have had destruction of certain parts of the brain. These studies have shown the following:

Removal of the *ideomotor cortex* makes is impossible for the person to conceive purposeful activities but does not prevent highly coordinated movements.

Removal of a small area of *motor cortex* causes loss of discrete movements in a given area of the body but not loss of many coordinate movements in the same area.

Removal of a small area of *premotor cortex* causes loss of some coordinate movements and *simple* patterns of movement in a specific area of the body but not loss of discrete movements in the same area.

Removal of the *somesthetic cortex* and *somesthetic association areas* makes is impossible for a person to perform *complex learned patterns* of movement but does not prevent his *attempting* to perform purposeful acts; nor does it prevent his performing *simple* patterns of movement.

Serious damage in the brain stem, particularly in the *reticular substance of the mesencephalon* or in the *thalamus,* can cause serious disruption of almost any type of motor activity.

On the basis of these studies, the following theory of motor control has been formulated: The thought that initiates motor activity begins in the

common integrative area of the cortex. Then, the activities to be performed are determined in the *ideomotor cortex,* operating in association with the common integrative area.

These areas do not control every single muscle movement. Instead, they "command" the motor and sensory regions of the brain to perform a particular activity that has been performed before and the memory of which is stored *in the somesthetic cortex and somesthetic association area,* the *sensory portion of the brain.* The motor cortex sends impulses to the muscles which begin the movement, and, simultaneously, sensory signals from the moving parts of the body are transmitted back to the sensory cortex, which determines whether the movement is going in the proper direction. If it is not, *corrective signals* are sent from the sensory cortex to the motor cortex to make the muscles move properly. Thus, the pattern in the sensory region controls the motor cortex which acts as a *servo system* to make the muscles "follow" this pattern.

However, some activities must be performed so rapidly that feedback impulses cannot return all the way from the peripheral parts of the body to the sensory cortex in time to effect proper control. In these instances, uncomplicated patterns of movement are believed to be stored in the premotor cortex, and these, so the theory goes, can be called forth without invoking the feedback control system of the sensory cortex.

FUNCTIONS OF LOWER CENTERS. Figure 236 illustrates diagrammatically the interrelationships of the different parts of the brain in the control of motor functions. This figure shows that lower centers as well as the cortex occupy a central position in motor control. The different parts of the cortex communicate with each other mainly through the thalamus and mesencephalon, so that the sensory cortex probably initiates most motor activities by stimulating lower centers of the brain which in turn excite the motor and premotor cortex. Also, many of the gross body movements and primitive movements, such as movements of the eyes, rotation of the head, and movements of equilibrium, can be controlled directly by lower centers in the mesencephalon, pons, and medulla without participation of the cortex at all.

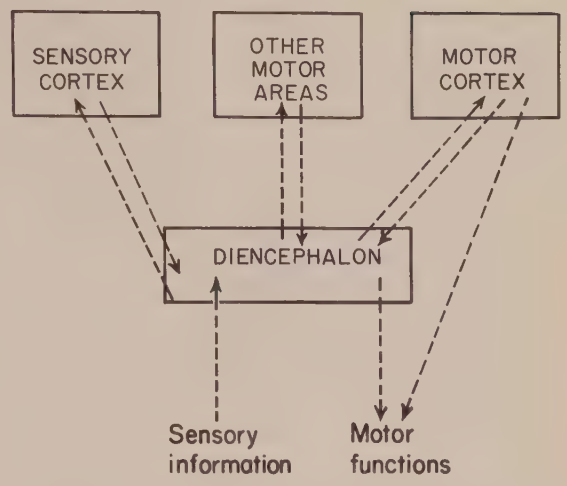


Figure 236. Principal pathways for performance of motor functions.

TRANSMISSION OF MOTOR SIGNALS TO THE SPINAL CORD AND MUSCLES

The Corticospinal Tract

Figure 237 illustrates the *corticospinal* or *pyramidal tract,* which transmits impulses from the motor cortex downward through the basal regions

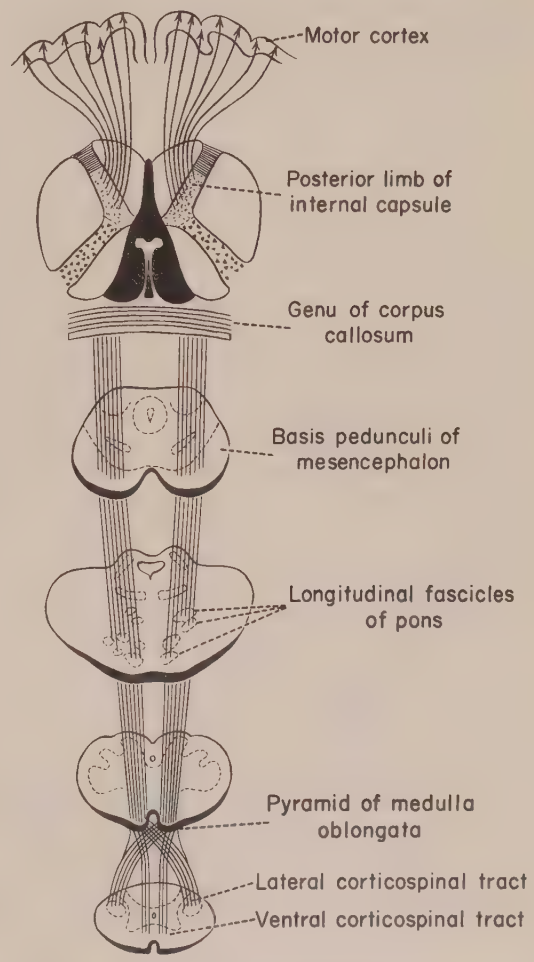


Figure 237. The corticospinal tract (also called the pyramidal tract) through which motor signals are transmitted from the motor cortex to the spinal cord. (Modified from Ranson and Clark: *The Anatomy of the Nervous System.* 10th Ed.)

of the brain, into the medulla, and finally into the spinal cord. Most of the fibers in the corticospinal tract cross to the opposite side in the medulla, and almost all of these which do not cross here do so in the upper segments of the spinal cord. As a result, stimulation of the left motor cortex contracts the muscles on the right side of the body, and, vice versa, stimulation of the right motor cortex contracts the muscles on the left side of the body.

The corticospinal tract is made up of two different types of nerve fibers. One type comes from large neurons in the motor cortex called *Betz cells* or *pyramidal cells*. Each motor cortex contains approximately 30,000 of these cells, and each one of the cells transmits signals to a discrete muscle, causing it to contract. The second group of fibers in the corticospinal tract is composed of very small fibers that originate in many small neurons of the motor cortex. These do not have the ability by themselves to stimulate the muscles but do have the ability to facilitate the neurons in the spinal cord that control the muscles. This facilitation makes these cells more active than usual, and, therefore, makes them rapidly responsive to signals transmitted by the Betz cells.

The Internuncial Cells of the Cord

Figure 238 illustrates a cross-section of the spinal cord, showing several of the fiber tracts that pass downward from the brain. The *corticospinal tract* is in the lateral, posterior portion of the cord. Nerve fibers from this tract terminate on *internuncial cells* located in the cord *gray matter*. Motor signals are usually transmitted through several additional internuncial cells before finally reaching the anterior motor neurons that control the muscles. The internuncial cells may be arranged in reverberating or parallel circuits for causing prolonged discharges, divergent circuits for amplifying the signal, or convergent circuits for integrating signals from several sources.

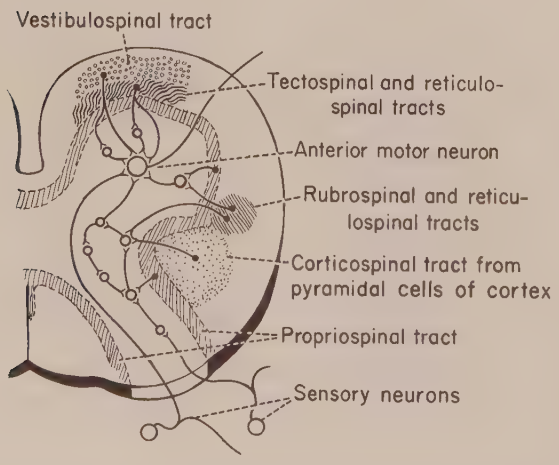


Figure 238. Convergence of all the different motor pathways on the anterior motor neuron.

A number of other fiber pathways besides those of the corticospinal tract terminate on the internuncial cells. These include *propriospinal fibers* that originate in other portions of the spinal cord, *sensory fibers* from the surface of the body, and fibers from the brain transmitted through the *rubrospinal tract*, the *reticulospinal tract*, and the *tectospinal tract*. Thus, the internuncial cells are not controlled entirely by the corticospinal tract, but instead by at least six or eight different types of fibers arriving from different sources in the nervous system and also from the peripheral nerves. These other fibers are mainly responsible for the automatic and subconscious movements of the muscles, in contrast to the conscious movements controlled by the corticospinal tract.

Some signals entering the internuncial cells are inhibitory rather than excitatory. These include especially signals from the reticulospinal tracts and from the vestibulospinal tracts.

The function of the internuncial cells is to integrate the signals from all possible sources and then to determine the type of motor signal to be directed to the anterior motor neurons. For instance, if a strong signal enters the internuncial cells from a sensory nerve at the same time that a signal enters from the corticospinal tract, the resultant stimulation of the anterior motor neuron will be much greater than will occur when a signal comes only from the corticospinal tract.

The Anterior Motor Neurons

The final pathway for transmission of motor signals from the brain to the muscles is through the *anterior motor neurons* which are very large neurons located in the anterior horns of the cord gray matter. On receiving impulses from the internuncial cells, these cells subtract the inhibitory signals from the excitatory signals, and the result determines the rate at which impulses will be transmitted into the motor nerve.

Fibers from the anterior motor neurons pass through the spinal nerves to all the skeletal muscles. Each fiber divides an average of approximately 150 times, and each of the terminal fibrils innervates one neuromyal junction on a muscle fiber. Therefore, each anterior motor neuron controls approximately 150 muscle fibers which constitute the motor unit described in Chapter 21.

Summary of Motor Transmission from the Brain to the Muscles

The discrete signals for muscular control originate in the Betz cells of the motor cortex. These signals are transmitted through the corticospinal tract into the spinal cord where they excite the internuncial cells of the cord gray matter. Here the signals are integrated with signals from other regions of the brain, and from peripheral nerves, to determine the overall degree of activity. If signals are arriving simultaneously from the corticospinal tract and other tracts, such as from the rubrospinal, tectospinal, and vestibulospinal tracts, the degree of activity will be much greater than if signals are arriving only through the corticospinal tract. Likewise, signals entering from the spinal nerves help to determine the degree of activity. The integrated signals from the internuncial cells then pass to the anterior motor neurons and from these to the muscles.

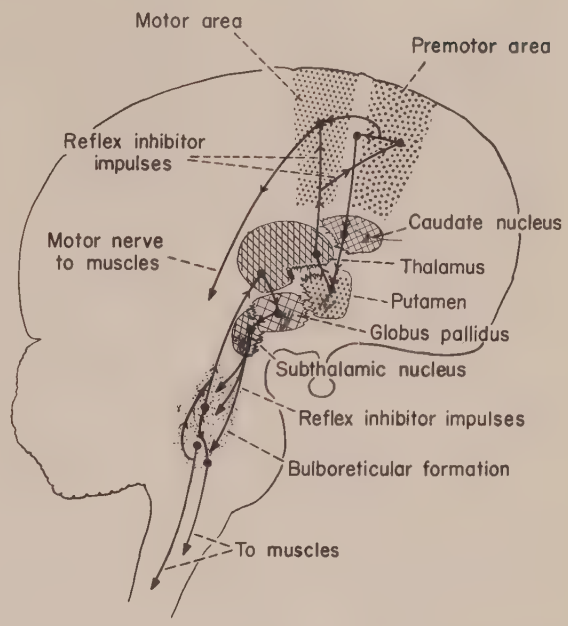


Figure 239. The basal ganglia and their connections with the cerebral cortex and other structures in the brain stem.

FUNCTION OF THE BASAL GANGLIA

The basal ganglia are composed of large masses of neurons located deep in the substance of the cerebrum and in the upper part of the mesencephalon. These are shown in Figure 239. They include the *caudate nucleus,* the *putamen,* the *globus pallidus,* the *subthalamic nucleus,* the *substantia nigra,* and several other less important nuclei.

Control of subconscious movements by the basal ganglia. Before attempting to discuss the function of the basal ganglia in man, we should speak briefly of their better known functions in lower animals. In birds, for instance, the cerebral cortex is very poorly developed, while the basal ganglia are highly developed. These ganglia perform essentially all the motor functions, even controlling the voluntary movements in much the same manner that the motor cortex of the human being controls voluntary movements. In the cat, and to a less extent in the dog, removal of the cerebral cortex prevents discrete types of motor functions but does not interfere with the cat's ability to walk, to eat perfectly well, to fight, to develop rage, to have periodic sleep and wakefulness, and even to participate naturally in sexual activities. However, if a major portion of the basal ganglia is destroyed, only gross stereotype movements remain, which are controlled by the very primitive lower areas in the brain stem.

In the human being, many of the potential functions of the basal ganglia are suppressed by the cerebral cortex, but if the cerebral cortex becomes destroyed in a very young human being, many of the voluntary motor functions do develop. The person will never be able to develop very discrete movements, particularly of the hands, but he can learn to walk, to control his equilibrium, to eat, to rotate his head, to perform almost any type of postural movement, and to carry out other subconscious movements. On the other hand, destruction of a major portion of the caudate nucleus almost totally paralyzes the opposite side of the body except for a few stereotyped reflex movements integrated in the cord or lower brain stem.

Unfortunately, little is known about the function of most of the individual basal ganglia other than the fact that they operate together in a loosely knit unit to perform the subconscious movements. Yet the following are a few of the discrete bits of information that are known about their individual functions.

(1) The *caudate nucleus* controls gross intentional movements of the body, these occurring mainly subconsciously but aiding in the overall control of body movements.

(2) The *putamen* operates in conjunction with the caudate nucleus to control gross intentional movements. Both of these nuclei also function in cooperation with the motor cortex to control many of the patterns of movement.

(3) The *globus pallidus* probably controls the "background" muscle tone for intended move-

ments. That is, if a person wishes to perform a very exact function with one of his hands, he first positions his body appropriately and then tenses the muscles of the upper arm. These functions are initiated by the globus pallidus.

(4) The *subthalamic nucleus* controls walking movements and perhaps other types of gross rhythmic body motions.

Abnormalities associated with damage of the basal ganglia. Even though we do not know all the precise functions of the basal ganglia, we do know many abnormalities that develop when portions of the basal ganglia are destroyed, as follows:

CHOREA. Chorea is random, uncontrolled contractions of different muscle groups of the body occurring continuously. Normal progression of movements cannot occur. Instead, the person may perform a normal sequence of movements for a few seconds and then suddenly jump to a new sequence, this occurring again and again without stopping. This is caused by widespread damage in the *caudate nucleus* and *putamen.*

ATHETOSIS. Athetosis is characterized by slow, writhing movements of peripheral parts of the body. For instance, the hand or arm may undergo worm-like movements, such as twisting of the arm to one side, then to the other side, then backward, then forward, repeating the same activity over and over again. The damage is always in the *globus pallidus.*

HEMIBALLISMUS. Hemiballismus is an uncontrollable succession of violent movements of large areas of the body. For instance, a leg may suddenly kick forward, and this may be repeated once every few seconds, or sometimes only once in many minutes. The damage that causes this is in the *subthalamus.*

PARKINSON'S DISEASE. This disease is characterized by *tremor* and *rigidity* of the musculature in either widespread or isolated areas of the body. The typical person with full-blown Parkinson's disease walks in a crouch like an ape, except that his muscles are obviously tense, his face is mask-like and he jerks all over with a violent tremor approximately 6 to 8 times per second. Yet when he attempts to perform voluntary movements, the tremor stops temporarily. This disease is caused by destruction of the *substantia nigra,* one of the less conspicuous basal ganglia that extends downward into the mesencephalon.

COORDINATION OF MOTOR MOVEMENTS BY THE CEREBELLUM

Movements of parts of the body are affected greatly by their inertia and momentum. That is, a limb requires a certain force to start it moving, but once started it keeps on moving until an opposing force stops the motion. The cerebral cortex is not organized to take these physical factors into consideration. Instead, the *cerebellum* makes the automatic adjustments that keep these factors from distorting the patterns of activity.

The cerebellum, shown in Figure 240, is a large structure located posterior to the brain stem. It receives signals from the proprioceptive receptors located in all joints, in all muscles, in the pressure areas of the body, and anywhere else that signals informing of the physical state of the body can be

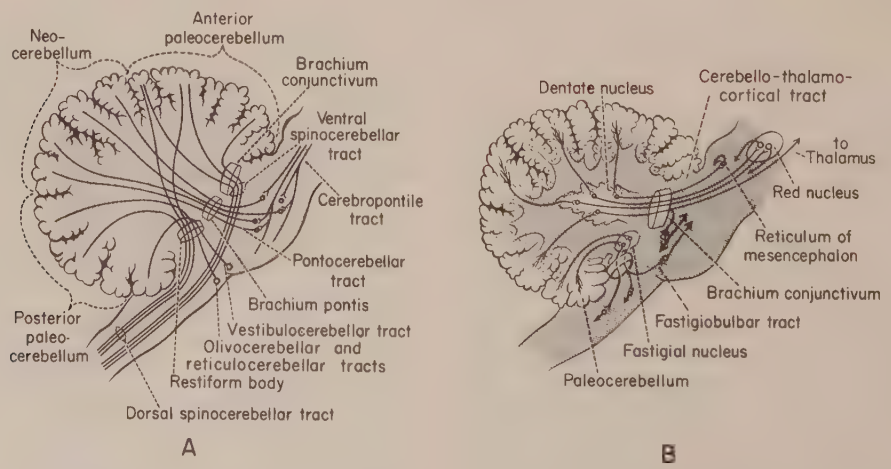


Figure 240. (A) Incoming fiber pathways to the cerebellum. (B) Outgoing fiber pathways from the cerebellum.

obtained. Signals are transmitted into the cerebellum, too, from the equilibrium apparatus of the ear, and even from the eyes to depict the visual relationship of the body to its surroundings. Finally, the cerebellum receives information directly from the motor and premotor cortex of all motor signals that are being sent to the muscles. In summary, the cerebellum is a collecting house for all possible information on the instantaneous physical status of the body.

Figure 240A illustrates the fiber tracts that transmit the various types of information into the cerebellum. These include (1) the *spinocerebellar tracts* that carry proprioceptor information from the body, (2) the *cortico-ponto-cerebellar tract* that carries motor information from the cerebrum to the cerebellum, and (3) the *vestibulocerebellar tract* that carries impulses from the equilibrium apparatus of the ear.

Once the cerebellum has collated its information on the physical status of the body, it transmits its analysis to other areas of the brain through the fiber tracts shown in Figure 240B, including especially (1) the *fastigiobulbar tracts* that go from the cerebellum into many of the structures of the brain stem, and (2) the *cerebello-thalamocortical tracts* that pass to the thalamus and then to the premotor and motor regions of the cerebral cortex.

The cerebellum acts as a feedback mechanism to help control the functions of the motor and premotor cortex; this is illustrated in Figure 241. It receives information from the cortex of the muscular movements that it is attempting to perform, while simultaneously receiving proprioceptive information directly from the body. In return, the cerebellum transmits appropriate signals back to the motor cortex and premotor cortex to apprise them of the momentary positions, rates of motion, and predicted changes in the status of the different parts of the body. Only when all this information is available can the motor cortex act intelligently to control the body movements. Some of the specific functions of the cerebellum are the following:

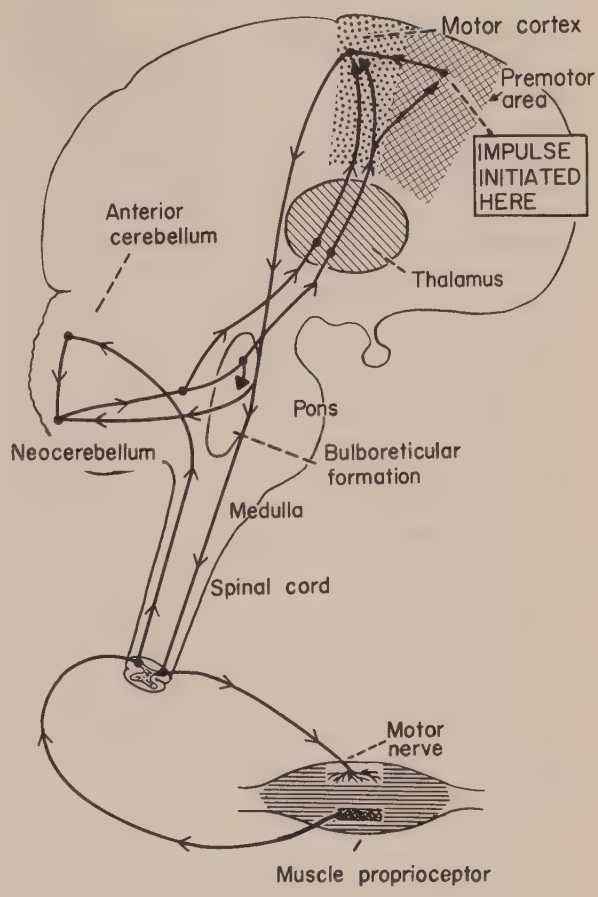


Figure 241. Feedback circuits of the cerebellum for damping motor movements.

Damping function of the cerebellum. When one moves his hand rapidly to a new position he can normally stop it exactly at the desired point, but without the cerebellum this cannot be accomplished. If the cerebellum is removed, the momentum of the arm carries the hand beyond the projected point until some other part of the brain can detect the fact that the hand has gone too far. The eyes and the proprioceptor mechanisms of the somesthetic cortex finally do detect the overshoot, and they set into play opposing muscular forces to bring the hand back to the projected point. Again momentum is built up during this return motion, and the hand moves too far in the opposite direction. Once more the cerebral mechanisms detect the overshoot and bring the hand back to approach the point. Gradually the successive overshoots become less and less until the hand at last reaches the desired point.

To prevent this effect of momentum, the cerebellum collects information from the moving parts of the body while they are actually moving and determines how much momentum is affecting the movements. Then, even before each part reaches its destination, the cerebellum sends "feedback" signals to the motor cortex to initiate appropriate "braking" contractions of opposing muscles to slow up and stop the movement at the proper point. In this way the finger can be brought to rest at a desired position without the annoying overshoot. This overall mechanism to prevent overshoot is called the *damping* function of the cerebellum.

ATAXIA. Lack of the damping function of the cerebellum causes the condition called *ataxia*, which means incoordinate contraction of the dif-

ferent muscles. For example, if a person who has lost his cerebellum tries to run, his feet will overshoot the necessary points on the ground for maintenance of equilibrium, and he will fall immediately. Even when walking, his gait will be very severely affected, for placement of the feet can never be precise; he falls first to one side, then over-corrects to the other side, and must correct again and again, giving him a broken gait. Ataxia in the hands may be so severe that it becomes impossible for the person to write, to nail a nail, or to perform any other precise movements without overshooting.

Predictive function of the cerebellum. Another function of the cerebellum closely allied to the damping function is its ability to predict the position of the different parts of the body ahead of time. When the leg is moving very rapidly forward while a person is running, the cerebellum, operating in conjunction with the somesthetic cortex, predicts where the leg will be at each instant during the next few hundredths of a second. It is because of this prediction that the person can send appropriate signals to the leg muscles, directing the exact point on the ground where the foot is to be placed to keep him from falling to one side or the other.

The predictive function also applies to the relationship of the body to surrounding objects, for without a cerebellum a person running toward a wall cannot predict when he will likely reach the wall. Monkeys that have had portions of the cerebellum destroyed have been known to run so rapidly toward walls, being unable to predict how rapidly they are approaching, that they bash their brains.

Equilibrium function of the cerebellum. A certain portion of the cerebellum, called the *flocculonodular lobes,* is concerned particularly with equilibrium of the body—that is, maintenance of the body in the upright position despite the pull of gravity. When the person tends to fall to one side, information to this effect enters the cerebellum from the equilibrium mechanism of the ear, from the visual sensations of the eyes, and from many of the proprioceptors of the body. This information is all summed together, and the neccessary signals to readjust the equilibrium are sent into the reticular substance of the brain stem and from there to the postural muscles. This is an entirely subconscious process, for the signals, instead of going through the cerebral cortex, are relayed directly from the brain stem to the muscles. The equilibrium mechanism will be considered in greater detail in the following chapter, which deals with cord and brain stem functions.

THE CONTROL OF SPEECH

The major characteristic of human beings that sets them apart from other animals is their ability

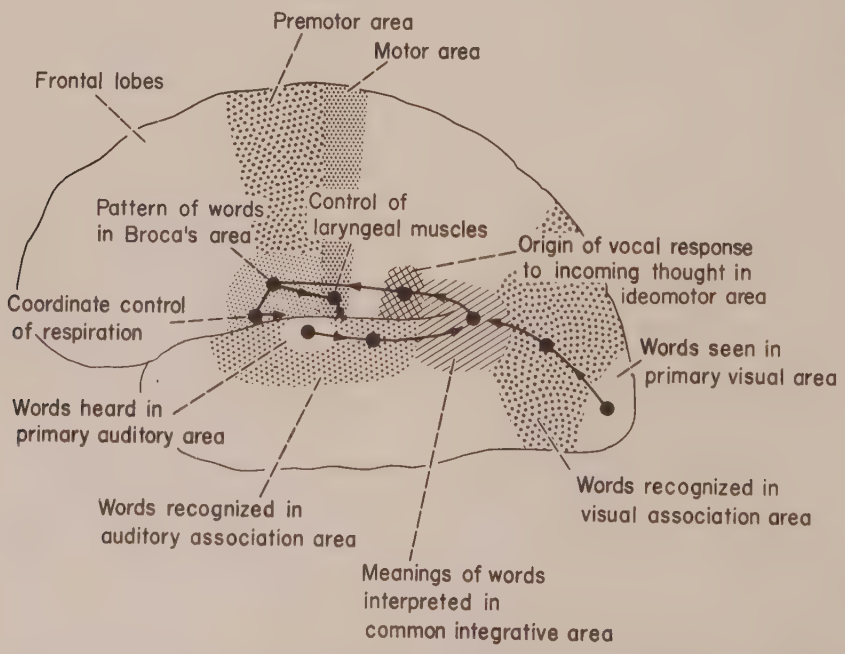


Figure 242. Pathways in the brain for communication, showing reception of thoughts by the auditory and visual pathways and control of speech by motor pathways.

to communicate one with the other. This depends on two highly developed functions of the brain: first, the ability to interpret speech and, second, the ability to translate thought into speech. These communicative functions require the highest degree of operative perfection in almost all parts of the brain. Therefore, a description of the act of communication provides an overall review of the integrative functions of the central nervous system.

Interpretation of communicated ideas. Ideas are generally communicated from one person to another either by sounds or written words. In the case of sounds, the information enters the primary auditory cortex in the superior part of the temporal lobe, as shown in Figure 242. The sounds then are interpreted as words and the words as sentences in the auditory association areas. The sentences are interpreted as thoughts in the common integrative region. Similarly, combinations of letters seen by the eyes are interpreted as words and the words as sentences by the visual association areas, and the sentences become thoughts in the common integrative region.

Motor functions of speech. In response to signals from the common integrative region or from other regions of the brain, the ideomotor area develops thoughts that it wishes to communicate to someone else. To do this it initiates a sequence of impulses, each impulse representing perhaps a syllable or a whole word, and transmits these into the premotor area that controls the larynx and mouth. This area is called *Broca's area,* or it is sometimes called simply the *speech center.* It is actually no more a speech center than the ideomotor area, but it is here that the muscular patterns for forming different sounds by the larynx and mouth are controlled. Impulses arriving from the ideomotor area set off a sequence of patterns in the speech center, and these in turn form the words. In addition to controlling the larynx and mouth, Broca's area sends impulses into an allied region of the premotor cortex that controls respiration. At the same time that the laryngeal and mouth movements occur, the respiratory muscles are contracted to provide appropriate air flow for the speech process.

REFERENCES

Brookhart, J. M.: The cerebellum. *Handbook of Physiology* 2:1245, 1960.

Cajal, S. R.: *Studies on the Cerebral Cortex.* Chicago, Year Book Publishers, Inc., 1955.

French, J. D.: The reticular formation. *Handbook of Physiology* 2:1281, 1960.

Greene, M.: *The Voice and Its Disorders.* New York, The Macmillan Company, 1957.

Jung, R., and Hassler, R.: The extrapyramidal motor system. *Handbook of Physiology* 2:863, 1960.

Neff, W. D., and Goldberg, J. M.: Higher functions of the central nervous system. *Ann. Rev. Physiol.* 22:499, 1960.

Olds, J.: Hypothalamic substrates of reward. *Physiol. Rev.* 42:554, 1962.

Patton, H. D., and Amassian, V. E.: The pyramidal tract: its excitation and functions. *Handbook of Physiology* 2:837, 1960.

Schaltenbrand, G., and Bailey, P.: *Introduction to Stereotaxis with an Atlas of the Human Brain.* New York, Grune & Stratton, Inc., 1958.

Terzuolo, C. A., and Adey, W. R.: Sensorimotor cortical activities. *Handbook of Physiology* 2:707, 1960.

REFLEX FUNCTIONS OF THE SPINAL CORD AND BRAIN STEM

25

Many of the functions of the spinal cord and hindbrain have already been discussed in the chapters on the sensory and motor systems of the central nervous system. The purpose of considering these areas now is to describe their automatic reflex functions. These aid in the overall performance of the sensory and motor activities, though they occur involuntarily.

Physiologic Anatomy of the Spinal Cord

Figure 243A illustrates a cross-section of the spinal cord, showing that it is composed of two major portions, the *white* and the *gray matter.* The gray matter, which lies deep inside the cord, has the appearance of double horns protruding anteriorly and posteriorly. The cell bodies of the neurons of the cord are located in the gray matter. The white matter, which comprises all other portions of the cord, is composed of fiber tracts. Several long *descending tracts* originate in the brain and pass down the cord to terminate on the neurons in the gray matter, and several other long tracts, the *ascending tracts,* originate in the cord and then pass upward to the brain.

In addition to the long descending and ascending tracts, many fibers called *propriospinal fibers* pass from one region of the cord to another, as shown in Figure 243B. This figure also illustrates

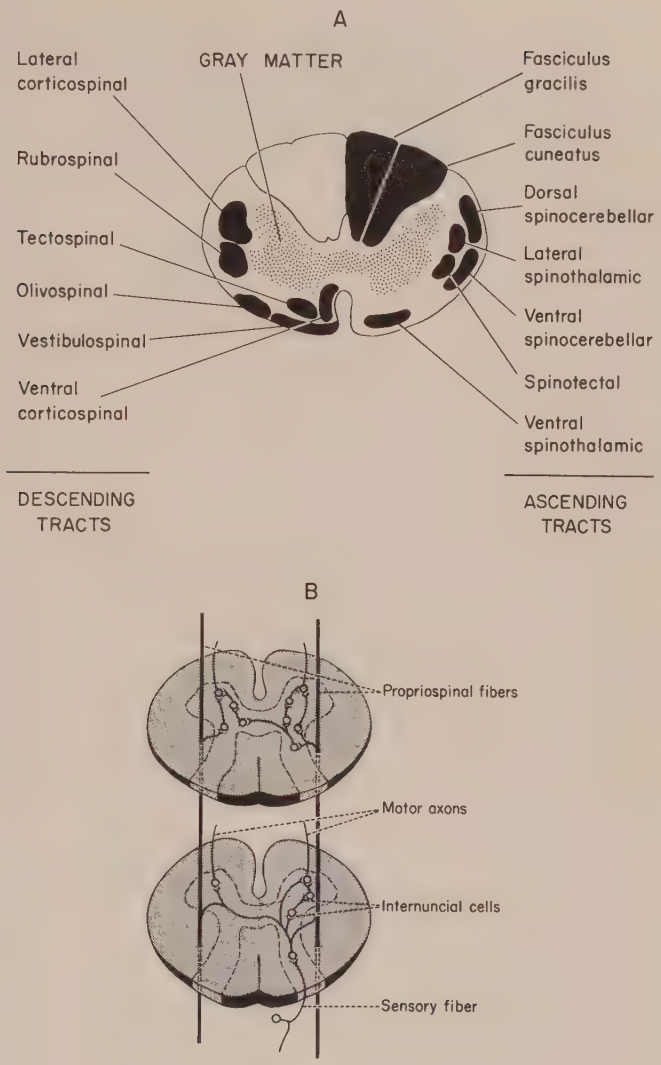


Figure 243. (A) Fiber tracts and gray matter of the spinal cord. (B) Two segments of the cord, illustrating the internuncial mechanisms that integrate cord reflexes and the propriospinal fibers that interconnect the cord segments.

the general pattern of organization of each segment of the spinal cord. Sensory nerve fibers enter the posterior horns of the gray matter on each side; some fibers synapse with neurons here while others pass directly up the cord. The neurons in turn send fibers in all directions through the cord. Some of these fibers enter the propriospinal tracts and pass upward or downward to adjacent segments, others pass to the internuncial cells, and a few pass directly to the anterior motor neurons. The anterior motor neurons then send fibers from the cord into the spinal nerves.

THE SIMPLE CORD REFLEXES

A neurogenic reflex is the transmission through nervous pathways of a signal initiated somewhere in the body to cause a reaction somewhere else in the body. Almost the entire nervous system functions as one great mass of reflexes. For example, the passage of somesthetic sensations from the skin to the cord or brain to cause some muscular reaction is a reflex.

Two essentials must always be present for a reflex to occur: a *receptor* organ and an *effector* organ. All of the sensory nerve receptors are receptor organs for reflexes, and all the muscles of the body, whether they be smooth muscles or skeletal muscles, are effector organs. Glandular cells that can be stimulated by nerve impulses are also effector organs.

The Axon Reflex

The simplest neuronal reflex is the axon reflex shown to the left in Figure 244; this reflex involves

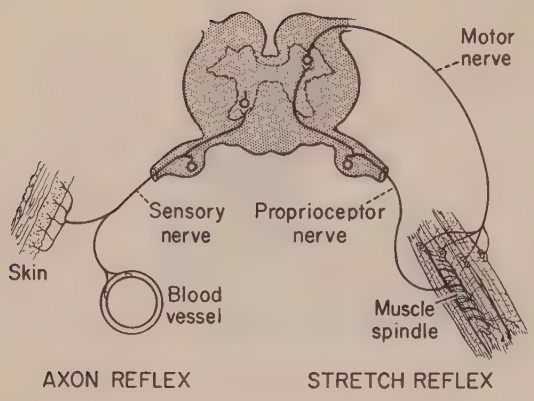


Figure 244. The neuronal mechanisms of the axon and stretch reflexes.

only part of a single neuron. Each sensory nerve fiber in a spinal nerve normally has a hundred or more branches, some of which terminate in the skin while others terminate on blood vessels or on other structures beneath the skin. It will be recalled that regardless of what point of a neuron is stimulated, the nerve impulse travels over the entire membrane of the fiber. Therefore, stimulating a single sensory receptor in the skin causes an impulse to travel into all the branches of the fiber as well as upward to the cord. When the impulse reaches the terminals near blood vessels, a hormone, probably histamine, is secreted, and this causes the blood vessels to dilate.

The axon reflex occurs principally when the skin is damaged. For example, scratching the skin with a pin elicits an axon reflex, which in turn causes blood to flow more rapidly than usual. The skin becomes reddened for several millimeters on each side of the scratch mark. In this way the axon reflex provides increased blood flow to damaged tissues and aids in their repair.

The Proprioceptor Reflexes

The spinal cord is the neurogenic locus of many different reflexes that originate in the proprioceptor receptors of the muscles, tendons, and other deep tissues. Signals from the receptors pass into the gray matter of the cord and cause motor impulses to be transmitted back to the muscles to cause appropriate muscular responses. Some of the proprioceptor reflexes are the following:

The stretch reflex. Sudden stretch of a muscle causes a reflex that makes the muscle contract. This reflex, which is illustrated to the right in Figure 244, is one of the simplest of all cord reflexes, for it involves only a single sensory neuron and a single motor neuron. Sudden stretch of the muscle stimulates muscle spindles in the muscle, and impulses from these excite the anterior motor neurons that control the same muscle. These, in turn, transmit reflex signals back to the muscle to cause it to contract.

THE KNEE JERK AND OTHER MUSCLE REFLEXES. One of the best known examples of the stretch reflex is the *knee jerk*, shown in Figure 245. Almost every doctor elicits this reflex when he performs a physical examination. A small hammer is used to strike the patellar tendon immediately below the kneecap. The sudden jolt to the tendon stretches the quadriceps muscle which in turn stretches muscle spindles and sends a barrage of impulses into the spinal cord. Then a reflex signal passes back to the quadriceps muscle, causing a sudden contraction that makes the lower leg jerk forward.

A muscle jerk similar to the knee jerk can be elicited in any muscle of the body by suddenly striking its tendon, or even by striking the muscle itself. The only essential for eliciting such a reflex is to stretch the muscle suddenly. Muscle reflexes of this type are elicited by physicans for two major purposes: First, if the reflex can be demonstrated, it is certain that both the sensory and motor nerve connections are intact between the muscle and the spinal cord. Second, muscle reflexes can help determine the degree of excitability of the spinal cord. When a large number of facilitatory impulses are being transmitted from the brain to the cord the muscle reflexes will be so active at times that simply tapping the patellar tendon with the tip of one's finger might make the leg jump a foot or more. On the other hand, the cord may be intensely inhibited by other impulses from the brain, in which case almost no degree of pounding on the muscles or tendons can elicit a response.

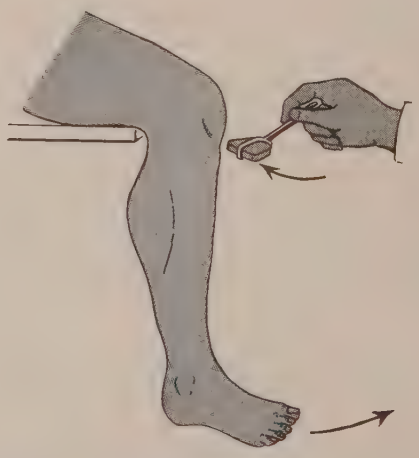


Figure 245. Method for eliciting the knee jerk.

RELATIONSHIP OF THE STRETCH REFLEX TO MUSCLE TONE. Muscle tone is ordinarily maintained by impulses that arise continuously in the muscle spindles. These pass into the spinal cord to cause reflex contraction of the muscles, which keeps the muscles in a slightly taut state all of the time. When this basic mechanism for maintaining muscle tone is absent, a signal from the brain must be far stronger than normally to elicit a response.

SERVO-POSITIONING FUNCTION OF THE MUSCLE SPINDLE. Figure 246 illustrates the relation of the muscle spindle to the surrounding skeletal muscle fibers. The muscle spindle is composed of a receptor area in its center called the *annulospiral receptor,* and a *muscular area* composed of *intrafusal muscle fibers* at both ends. The intrafusal muscle fibers are controlled by a special set of nerve fibers from the spinal cord called *gamma efferent fibers.*

The muscle spindle operates to cause the muscle to contract to a particular length as follows: Impulses transmitted from the spinal cord through the gamma efferent nerves shorten the intrafusal muscle fibers. If the skeletal muscle fibers do not shorten simultaneously, the central portion of the muscle spindle becomes stretched and excites the annulospiral receptors. Impulses are then transmitted into the spinal cord, causing reflex contraction of the skeletal muscle fibers. Thus, the original contraction of the intrafusal fibers causes automatic reflex contraction of the skeletal muscle fibers. In other words, the muscle *follows* the degree of contraction of the intrafusal fibers, which is called *servo-positioning* of the muscle. This type of contraction is used in the performance of postural movements or in positioning limbs to exact positions with respect to the rest of the body.

DAMPING FUNCTION OF THE STRETCH REFLEX. If a limb or some other part of the body is suddenly moved, an immediate stretch reflex is elicited which tends to oppose this movement and thereby attempts to maintain the status quo. This is called the *damping function* of the stretch reflex. This function also provides smooth motions rather than jerky motions in response to signals from the brain, because most signals from the brain arrive in the cord in sudden bursts rather than continuously.

The tendon reflex. A small, branching receptor that looks like a bush and is called the *Golgi tendon apparatus* is found in all muscle tendons. It detects the degree of stretch of the tendons, and its function is to protect the tendon and muscle from excessive stretch. The reflex that performs this function is the following: Ordinarily, light and moderate degrees of stretch cause no reflex whatsoever; however, if the degree of stretch on the tendon exceeds a certain critical value, an immediate reflex occurs to *inhibit* the anterior motor neurons that innervate the muscle. The muscle immediately relaxes, and the excessive stretch is removed from the muscle. Obviously, this is an important protective reflex to prevent damage to the muscle or tendon.

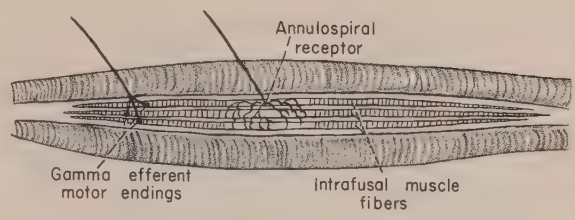


Figure 246. Arrangement of the muscle spindle with respect to the surrounding muscle fibers.

The extensor thrust reflex. An important proprioceptor reflex that helps support the body against gravity is the *extensor thrust reflex.* Pressure on the pads of the feet causes automatic tightening of the extensor muscles of the legs. This reflex is initiated by pressure receptors in the bottom of the foot. The signal passes to the internuncial cells in the cord. Here it is amplified and diverges into an appropriate pattern of impulses to tighten the extensor muscles, allowing the animal to keep his leg stiffened automatically when standing.

The magnet reaction. A proprioceptor reflex closely allied to the extensor thrust reflex, but still more complicated, is the so-called *magnet reaction.* In an animal that has had its spinal cord cut so that impulses from the brain will not interfere, one can place the tip of his finger on the pad of the animal's foot, then move his finger in all directions, and the foot will follow the finger. Moving the finger to one side causes appropriate proprioceptor reflexes to make the limb move in the direction of the force. This reaction is an aid to equilibrium of the animal, for excess pressure on one side of the foot indicates that he is falling in that direction, and automatic stiffening of the leg in that direction helps to prevent the fall.

The Flexor Reflex

Pain causes automatic withdrawal of any pained portion of the body from the object causing the pain. This is called a *flexor reflex* or sometimes a *pain reflex,* a *nociceptive reflex,* or a *withdrawal reflex.*

The neuronal mechanism of the flexor reflex is

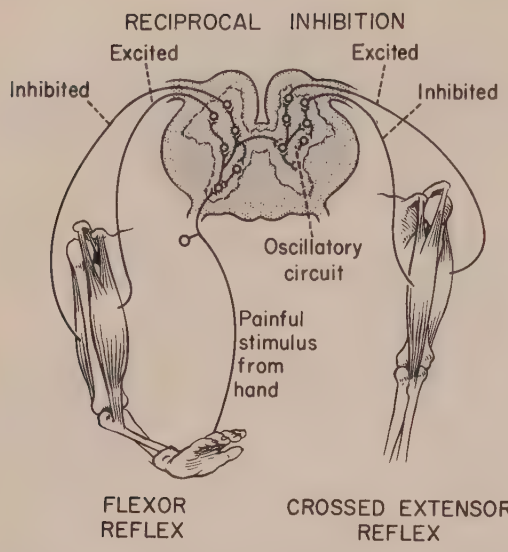


Figure 247. Neuronal mechanisms of the flexor reflex, of the crossed extensor reflex, and of reciprocal inhibition.

shown to the left in Figure 247. On entering the gray matter of the cord, the pain signal stimulates internuncial cells that transmit a pattern of impulses through appropriate anterior motor neurons to cause withdrawal of the area of the body that is being pained. In Figure 247 the pain stimulus is initiated in the hand, and the biceps muscle becomes excited and pulls the hand away.

The flexor reflex normally lasts for only a split second after the painful stimulus has been applied, but this short period of time is usually sufficient to pull away from the stimulus.

The Crossed Extensor Reflex

When a flexor reflex occurs in one limb, impulses also pass to the opposite side of the cord, where they stimulate the internuncial cells controlling the extensor muscles of the opposite limb, thus causing it to extend. This is called a *crossed extensor* reflex. For example, a painful stimulus applied to the left hand causes the right arm to extend, as shown in Figure 247. Such a reflex pushes the person away from the painful object. That is, at the same time that he withdraws his pained hand on the left, he pushes his entire body away on the right.

Withdrawal reflexes are present in all parts of the body, though organized somewhat differently in different areas. For instance, if a needle pricks the small of a person's back, he automatically arches forward, which causes withdrawal. Though this reflex is slightly different from the flexor and crossed extensor reflexes, its result is the same.

Reciprocal Inhibition

A very important feature of most reflexes, especially illustrated by the flexor and crossed extensor reflexes, is the phenomenon called *reciprocal inhibition.* That is, when a reflex excites a muscle, it ordinarily inhibits the antagonistic muscle at the same time. For example, when the flexor reflex of Figure 247 excites the biceps muscle, it simultaneously inhibits the opposing triceps muscle. Also, the crossed extensor reflex excites the triceps but inhibits the biceps. In all parts of the body where opposing muscles exist, a corresponding reciprocal inhibition circuit is present in the spinal cord. This obviously allows greater ease in the performance of desired activities.

RHYTHMIC REFLEXES OF THE CORD

The reflexes of the spinal cord are not limited to single reactions; frequently rhythmic reflexes also occur. That is, "patterns of activity" in the cord can make a portion of the body move back and forth in a rhythmic motion. Two especially important types of rhythmic reflexes controlled entirely by the spinal cord are the scratch reflex and the walking reflexes.

The Scratch Reflex

The scratch reflex, though not of importance in the human being, is one of the most important of all protective mechanisms in lower animals. A dog can actually scratch away a flea or other irritating object even after his spinal cord has been completely transected.

The scratch reflex depends on two separate integrative abilities of the spinal cord: First, the cord must provide the rhythmic to and fro motion of the leg. This is accomplished by reverberating circuits in the internuncial cells that send impulses during part of the cycle to one group of muscles, then to the opposing muscles, contracting alternately both sets of muscles to cause the rhythmic motion. The second cord mechanism utilized in the scratching act pinpoints the area of the body that needs scratching. For instance, a flea moving across the back of a sleeping dog is followed exactly by the scratching paw. When the flea crawls across the midpoint of the animal's belly this paw stops scratching, but the paw on the opposite side of the body immediately finds the flea and begins to scratch. Thus, the scratch reflex is one of the most complicated of all cord reflexes for it requires, first, localization of the

irritation, second, coordinate movement of the paw into position, and, third, reverberating motion of the limb.

Walking Reflexes

The spinal cord also is capable of providing the rhythmic to and fro walking movements of the legs. These movements, like the scratching act, are controlled by reverberating circuits in the internuncial cells that contract alternately the antagonistic pairs of muscles. Reciprocal inhibition also plays a part, for this keeps the muscles on the two sides of the body operating at opposite phase to each other, allowing one leg to move forward while the opposite one moves backward. In lower animals, fiber pathways also pass from the hindlimb region of the cord to the forelimb region, to cause appropriate phasing of the hind and forelimb movements. This cord control of walking movements with appropriate phasing of the limbs is illustrated by Figure 248, which shows a dog with his spinal cord transected at the neck. He is nevertheless still exhibiting continuous, rhythmic, to and fro movements of his limbs.

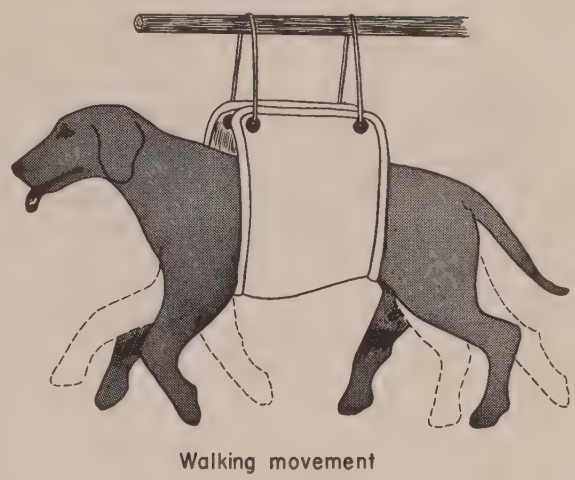


Figure 248. Walking movements in a dog whose spinal cord has been transected at the neck.

AUTONOMIC REFLEXES OF THE CORD

In addition to the cord reflexes for controlling skeletal muscular action, the spinal cord also harbors reflex circuits that help to control the visceral functions of the body. These reflexes are initiated by sensory receptors in the viscera, and the signals are transmitted through sensory nerves to the internuncial cells of the cord gray matter, where appropriate patterns of reflex responses are determined. Then the signals pass to *autonomic motor neurons* located in the cord gray matter about midway between the anterior and posterior horns. These cells send impulses into the sympathetic chain and back to the viscera to cause autonomic stimulation.

The Peritoneal Reflex

One of the most important of the autonomic reflexes is the peritoneal reflex. Whenever any portion of the peritoneum is damaged, this reflex slows up or stops all motor activity in the nearby viscera. For example, in appendicitis, movement of food through the gastrointestinal tract stops almost completely, preventing further irritation of the inflamed appendix and allowing the reparative processes of the body to function at optimum efficiency. Without this reflex many would die of appendicitis before aid could be given.

Vascular Control by the Cord

The blood pressure is normally controlled by vasomotor impulses transmitted from the brain, but after the spinal cord has been transected these impulses can no longer reach the blood vessels. Yet cord reflexes are still capable of modifying local vascular blood flow in response to such factors as pain, heat, and cold. For instance, when one of the viscera is pained tremendously by overdistention, such as overdistention of the urinary bladder, the blood pressure may rise to twice normal. Though these effects of cord reflexes on the blood pressure are not important normally, they do illustrate the capability of the cord to help control many of the involuntary reactions of the autonomic nervous system.

The Bladder and Rectal Reflexes

Among the most important of the autonomic reflexes of the cord are the bladder and rectal reflexes that cause automatic emptying of these two organs when they become filled. When either the bladder or the rectum becomes excessively full, sensory signals are transmitted into the internuncial cells of the lower end of the cord, as shown in Figure 249. Appropriate signals are then transmitted through parasympathetic nerves back to the bladder or colon. Here they excite the main body of the respective organ, but at the same

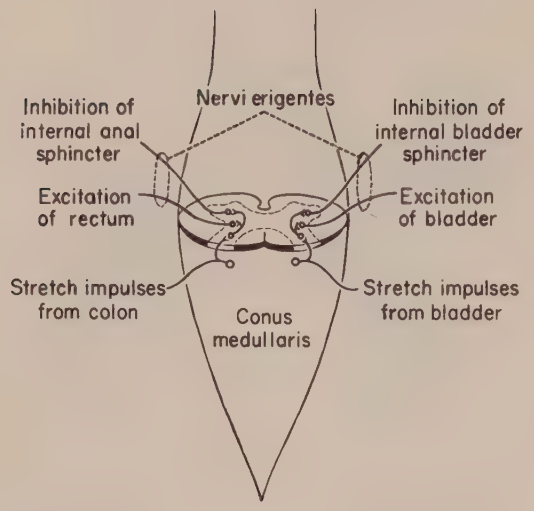


Figure 249. Neuronal circuits of the bladder- and rectal-emptying reflexes.

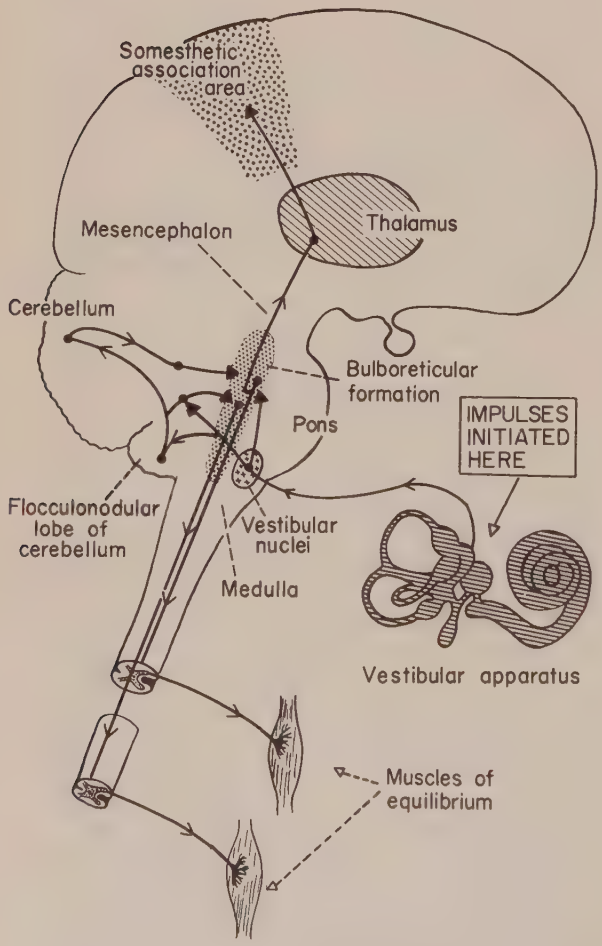


Figure 250. Nervous mechanisms of equilibrium.

time inhibit the internal sphincter of the urethra or anus, thereby causing emptying.

In persons whose spinal cords have been cut, these automatic reflexes are sometimes effective enough to keep the bladder and colon evacuated satisfactorily all the time. In the normal person, however, these reflexes are inhibited by impulses from the brain until an opportune time arises for evacuation. Unfortunately, this inhibition frequently leads to constipation or painful distention of the bladder.

MOTOR FUNCTIONS OF THE BRAIN STEM

The brain stem is composed of the medulla, the pons, and the mesencephalon. It is in these areas that many of the centers for blood pressure control, respiration, and gastrointestinal regulation are located, all of which are discussed in detail in other chapters. The purpose of the present section is not to give an overall discussion of the brain stem, but to present its motor functions and their relation to the cord and the cerebrum. The motor functions of the brain stem can be divided into two major parts: first, its function in helping the animal support his body against gravity, and, second, its function in the maintenance of equilibrium.

The bulboreticular formation of the brain stem. Figure 250 shows the *bulboreticular formation*, comprising most of the brain stem, and some of its connections with other parts of the body. Stimulation of this formation transmits impulses down the *reticulospinal, tectospinal, vestibulospinal,* and *rubrospinal tracts* into the cord to control the muscles throughout the body.

The bulboreticular area receives incoming fibers from many sources: fibers direct from all areas of the body, from the cerebellum, from the equilibrium apparatus of the ear, from the motor cortex, and from the basal ganglia. Thus, the bulboreticular formation is an integrative area for combining and coordinating (1) sensory information from the body, (2) motor information from the motor cortex, (3) equilibrium information from the vestibular apparatuses, and (4) proprioceptor information from the cerebellum. With this information available it controls many of the involuntary muscular activities.

Support of the Body against Gravity

Widespread stimulation of the upper bulboreticular formation excites mainly the extensor mus-

cles of the body, causing the trunk and the limbs to become stiffened, and making it possible to stand erect. Without this stiffening, the body would immediately crumple because of the pull of gravity.

Suppression of the bulboreticular area—decerebrate rigidity. When one wants to sit rather than to stand, the excitation of the muscles by the bulboreticular formation must be inhibited. This is accomplished mainly by suppressor impulses from the basal ganglia. When the basal ganglia are damaged, or when all the fibers tracts from the cerebrum are cut, the bulboreticular formation automatically becomes so active that the animal becomes rigid all over. This phenomenon, called *decerebrate rigidity,* is illustrated in the dog of Figure 251. It shows the necessity of the basal ganglia to inhibit the bulboreticular formation when motor functions other than the support of the body are to be performed.

Coordination of contraction of antagonistic postural muscles. Another function of the basal ganglia in relation to the bulboreticular formation is to help coordinate the contraction between antagonistic postural muscles. The degree of contraction of antagonistic muscles must be controlled appropriately at all times to keep one of the muscles from overbalancing the other. Signals transmitted from the bulboreticular formation up to the basal ganglia and then back again to the bulboreticular formation are probably responsible for this coordination. Without this circuit, the antagonistic muscles reverberate back and forth, trying to establish appropriate degrees of contraction but never being able to do so. This is probably the cause of the parkinsonian tremor which was described in the previous chapter.

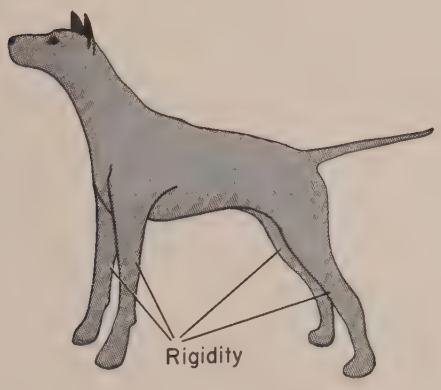


Figure 251. Rigidity in a dog whose brain has been transected above the bulboreticular formation.

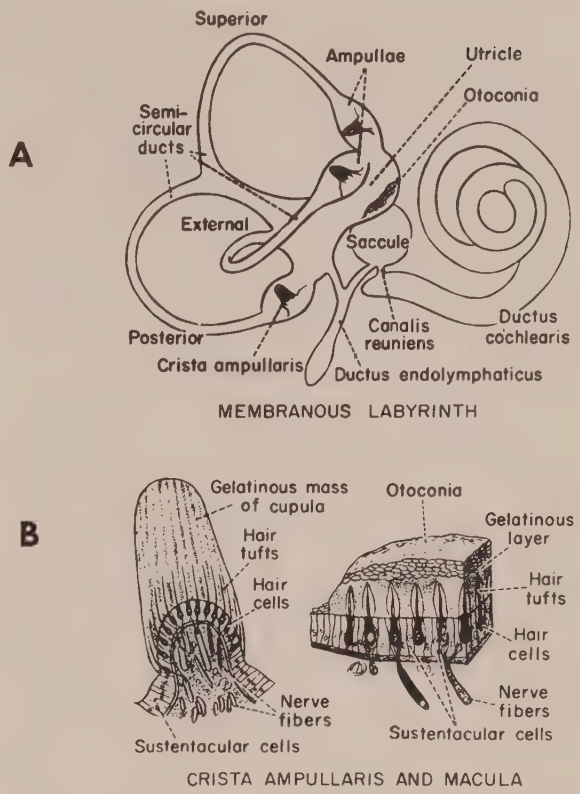


Figure 252. Functional structures of the vestibular apparatus.

Equilibrium Function of the Brain Stem

In addition to providing the neurogenic mechanisms for support of the body against gravity, the bulboreticular formation is also capable of maintaining equilibrium by varying the degree of tone in the different extensor muscles. Most of the equilibrium reflexes are initiated by the vestibular apparatuses located on each side of the head, adjacent to each internal ear.

Function of the vestibular apparatus. The vestibular apparatus and its connections with the equilibrium areas of the central nervous system are shown in Figure 252. This apparatus detects the position of the head in space—that is, it determines whether the head is upright with respect to the gravitational pull of the earth or whether it is lying back, upside down, or in another position. It also detects sudden changes in direction of movement. To perform these functions the vestibular apparatus is divided into two separate physiologic sections, the *utricle* and the *semicircular canals.*

THE UTRICLE. On the wall of the utricle is a structure called the *macula,* which is illustrated in Figure 252B. Nerve cells in the base of the macula project "hairs" upward into a gelatinous

mass in which are located many minute ossified granules, like very small granules of sand, called *otoconia.* When the head is bent to one side the weight of the otoconia pushes the hairs to that side, thereby stimulating the nerves. In this way the utricle supplies the equilibrium regions of the central nervous system with the information needed to maintain balance.

The utricle also helps the person maintain his balance when he suddenly begins to move forward, to one side, or in any other linear direction. When he begins to move forward, the inertia of the otoconia causes them to lag behind the movement of the rest of the body and, therefore, to bend the hairs backward. This gives the sensation of falling off balance in the backward direction. As a result, one leans forward to correct this imbalance, which explains why an athlete leans forward automatically when he first starts to run.

On the other hand, when one wishes to decelerate, he must lean backward. Again it is the otoconia of the utricle that initiate automatically this backward leaning, for when one brakes himself the momentum of the otoconia keeps them moving forward while the rest of the body slows. This bends the hairs of the macula in the forward direction, making the person feel as if he were falling head first toward the ground. As a result, the equilibrium mechanism leans the body backward automatically.

THE SEMICIRCULAR CANALS. The *semicircular canals* are small circular tubes that contain fluid, as shown in Figure 252A. The vestibular apparatus on each side of the head has three separate canals located respectively in the three planes of space, one in the horizontal and the other two in the two vertical planes. If one suddenly turns his head in any direction, the fluid in one or more of the semicircular canals lags behind the movement of the head because of the inertia of the fluid. This is the same effect as that observed when one suddenly rotates a glass of water; the glass rotates, but the water remains still. As the fluid moves in the semicircular canal it flows against the *crista ampullaris,* shown at the left in Figure 252B, which is a valve-like leaflet located at one end of the canal. This structure contains hair tufts from nerve cells like those in the macula, and bending these tufts to one side or the other gives a person the sensation that his head is beginning to turn.

The information transmitted from the semicircular canals apprises the nervous system of sudden changes in direction of movement. With this information available the bulboreticular formation can correct for any imbalance that is likely to occur *even before it does occur.* This is particularly important when one is changing direction of movement rapidly, as when playing a fast game of almost any type.

Function of the cerebellum in equilibrium. In addition to transmitting signals into the bulboreticular formation, the semicircular canals and utricle also send information to the *flocculonodular lobes* of the cerebellum. Since one of the major functions of the cerebellum is to predict future position of the body in space, the function of the flocculonodular lobes is probably to predict ahead of time when a state of imbalance is about to occur. This allows appropriate corrective signals to the bulboreticular formation even before the person falls off balance, and prevents imbalance from ever occurring rather than necessitating attempts to correct it after it has already occurred. Persons who lose their cerebellum lose this ability to predict ahead of time, and, as a result, must perform all movements slowly or else fall.

The neck proprioceptor receptors and their relationship to the vestibular mechanisms. The vestibular apparatuses detect only the position of the *head,* not of the *body,* in relation to the pull of gravity. To translate this information from the head to the whole body, the relationship of the head to the body must be known. This knowledge is provided by proprioceptor receptors in the neck. For instance, when the head is bent backward, the vestibular apparatuses send information that the position of the head is changing, but at the same time the neck proprioceptors send information that the head is angulating backward in relation to the rest of the body. The two sets of impulses negate each other and the tautness of the postural muscles remains exactly the same. In other words, the proprioceptor reflexes from the neck are as necessary for regulating equilibrium as the complicated reflexes initiated by the vestibular apparatuses.

Peripheral proprioceptor and visual mechanisms of equilibrium. After the vestibular apparatuses have been destroyed, a person can still maintain his equilibrium provided he moves slowly. This is accomplished mainly by means of proprioceptor information from the limbs and surfaces of the body and visual information from the eyes. If he begins to fall forward, the pressure on the anterior parts of his feet increases, stimulating the pressure receptors. This information transmitted to his brain helps to correct the imbalance. At the same time, his eyes also detect the lack of equilibrium, and this information too helps to correct the situation.

Unfortunately, the visual and proprioceptor systems for maintaining equilibrium are not

organized for rapid action, which explains why a person without his vestibular apparatuses must move slowly.

OVERALL CONTROL OF LOCOMOTION

From the foregoing discussions of the motor functions of the spinal cord, brain stem, and cerebrum, it is now possible to construct the overall pattern of locomotion. First, the person must support himself against gravity. This is accomplished partly by the extensor thrust mechanism of the spinal cord that allows stiffening of the limbs when pressure is applied to the pads of the feet; in addition to this, the bulboreticular formation transmits impulses continuously into the extensor muscles to keep the body and limbs stiff. Also, the degree of stiffening of the different parts of the body is varied by the equilibrium apparatus so that when one tends to fall the appropriate muscles are contracted to bring the body back into the upright position.

Once the body is supported against gravity and maintained in a state of equilibrium, locomotion then depends on rhythmic motion of the limbs. Rhythmic circuits in the spinal cord are capable of providing the to and fro movement of the limbs, and the movements of the opposing limbs are kept in opposite phase with each other by the reciprocal inhibition mechanisms of the cord. Thus, most of the functions of locomotion can be provided by the cord and brain stem, but the cerebral cortex must control these functions in accord with the desires of the individual. When he wishes to move forward, to stop, or to turn to one side, his motor cortex and basal ganglia simply initiate the action, stop it, or change it by sending *command signals*. The cord provides the stereotyped actions required to perform the actual movements. In this way, the energy of the portion of the brain operating at a conscious level is conserved to perform other mental feats.

REFERENCES

Bronisch, F. W.: *The Clinically Important Reflexes.* New York, Grune & Stratton, Inc., 1952.

Creed, R. S., Denny-Brown, D., Eccles, J. C., Liddell, E. G. T., and Sherrington, C. S.: *Reflex Activity of the Spinal Cord.* Oxford, Clarendon Press, 1932.

Denny-Brown, D.: Motor mechanisms—introduction: the general principles of motor integration. *Handbook of Physiology* 2:781, 1960.

Eldred, E.: Posture and locomotion. *Handbook of Physiology* 2:1067, 1960.

Fischer, J.: *The Labyrinth: Physiology and Functional Tests.* New York, Grune & Stratton, Inc., 1956.

Hausman, L.: *Atlases of the Spinal Cord and Brainstem and the Forebrain.* Springfield, Ill., Charles C Thomas, 1956.

Hunt, C. C., and Perl, E. R.: Spinal reflex mechanisms concerned with skeletal muscle. *Physiol. Rev.* 40:538, 1960.

Lloyd, D. P. C.: Spinal mechanisms involved in somatic activities. *Handbook of Physiology* 2:929, 1960.

THE AUTONOMIC NERVOUS SYSTEM, SLEEP, AND PSYCHOSOMATIC EFFECTS

26

The autonomic nervous system operates involuntarily to control many of the internal functions of the body. Various aspects of these have already been described in relation to bladder and rectal reflexes, peritoneal reflexes, the control of arterial pressure, and the control of many internal organs. The purpose of the present chapter is to present the overall function of the autonomic nervous system, pointing out its special attributes and peculiarities.

SYMPATHETIC AND PARASYMPATHETIC DIVISIONS OF THE AUTONOMIC NERVOUS SYSTEM

The autonomic nervous system is considered to have two separate divisions, the *sympathetics* and the *parasympathetics*. The reasons for this division are: First, the anatomic distributions of the nerve fibers in the two divisions are distinct from each other. Second, the effects of the two divisions on the organs are usually antagonistic to each other. Third, the types of hormones secreted at the nerve endings are usually different in the two systems.

Anatomy of the sympathetic nervous system. Figure 253 illustrates the anatomy of the sympathetic nervous system, showing the sympathetic chain lying on each side of the spinal cord, and the chain's connections with the cord and peripheral organs. The sympathetic chain is connected with the spinal cord in a rather peculiar manner, shown in Figure 254. Sympathetic fiber tracts leave the cord through the anterior roots of the spinal nerve. Then, after traveling less than a centimeter, they pass through a small whitish nerve called the *white ramus* into the *sympathetic chain.* From here, fibers travel in two directions. Some pass into visceral *sympathetic nerves* that innervate the internal organs of the body, while the others return through another small nerve called the *gray ramus* back into the spinal nerve. These latter fibers then travel all through the body along the spinal nerves to supply the blood vessels, the sweat glands, and even the pilo-erector muscles that cause the hairs to stand on end.

Sympathetic fibers enter the sympathetic chain only in the thoracic and upper lumbar regions, and none enter the chain in the neck, lower lumbar, or sacral regions. To supply the head with sympathetic innervation, sympathetic fibers from the thoracic chain extend upward into the neck and then to all the structures of the head. Also, sympathetic fibers pass downward from the chain into the lower abdomen and legs.

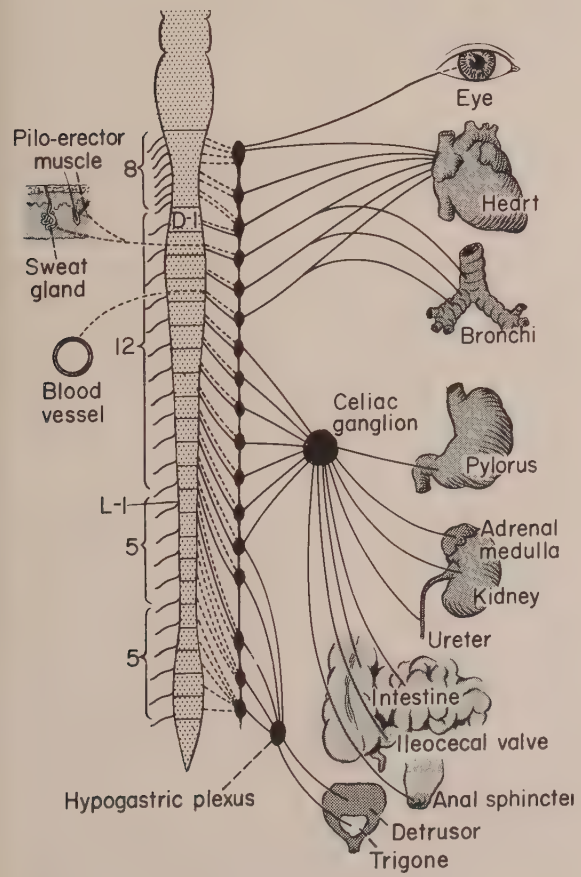


Figure 253. Anatomy of the sympathetic nervous system.

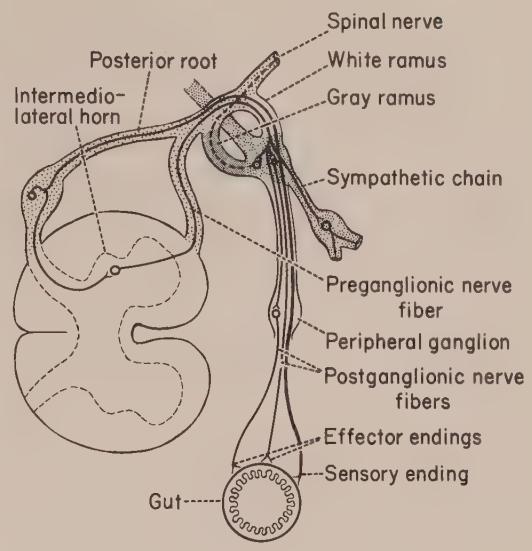


Figure 254. Connections between the spinal cord, a spinal nerve, and the sympathetic chain.

VISCERAL SENSORY FIBERS IN THE SYMPATHETIC NERVES. Sensory fibers pass through the sympathetic nerves along with the sympathetic fibers. These arise in the internal organs, then enter the sympathetic nerves, and finally travel by way of the white rami into the spinal nerves. From here they enter the posterior horns of the cord gray matter and either cause autonomic cord reflexes or transmit sensations to the brain in the same manner that sensations are transmitted from the surface of the body. These sensations are the visceral sensations that were considered in detail in Chapter 24.

PREGANGLIONIC AND POSTGANGLIONIC NEURONS OF THE SYMPATHETIC NERVOUS SYSTEM. Figure 253 shows bulbous enlargements located periodically along the sympathetic chain and at different points among the sympathetic nerves after they leave the chain. These enlargements are called *ganglia.*

Sympathetic signals are transmitted from the spinal cord to the periphery through two successive neurons. The cell body of the first neuron is located in the spinal cord in the lateral gray matter. The fiber from this neuron passes into the sympathetic system as illustrated in Figure 254. It synapses with a second neuron in either a chain or peripheral ganglion. The fiber from the second neuron passes directly to the organ to be con-

trolled. The first neuron, from the cord to the ganglion, is called the *preganglionic neuron;* the second, from the ganglion to the organ, is called the *postganglionic neuron.* Thus, the sympathetic motor system is different from the skeletal motor system, for skeletal muscles are stimulated through a single neuron rather than through two.

Anatomy of the parasympathetic nervous system. Figure 255 shows the parasympathetic nervous system. The fibers of this system originate mainly in the tenth cranial nerve, called the *vagus nerve.* However, a few fibers originate in the third, fifth, and seventh cranial nerves, and also in several of the sacral segments of the spinal cord. The vagus nerve supplies parasympathetic fibers to the heart, the lungs, and almost all of the organs of the abdomen. The other cranial nerves supply the head, and the sacral fibers supply the urinary bladder and the distal parts of the colon. However, since approximately 90 per cent of all the parasympathetic fibers of the body pass through the vagus nerve, most physiologists, in thinking about the parasympathetic system, think almost automatically of the vagus nerve itself.

PREGANGLIONIC AND POSTGANGLIONIC NEURONS OF THE PARASYMPATHETIC SYSTEM. The parasympathetic system is like the sympathetics in that impulses travel through preganglionic and postganglionic neurons. The cell bodies of the preganglionic neurons are in the brain stem or sacral cord, and their fibers usually pass all the way to the wall of the organ to be stimulated, where they synapse with postganglionic neurons. The postganglionic fibers then travel only a few millimeters, or at most a few centimeters, before they reach their final destination. This is different from the sympathetics, whose postganglionic cells are located in the sympathetic ganglia long distances from the respective organs.

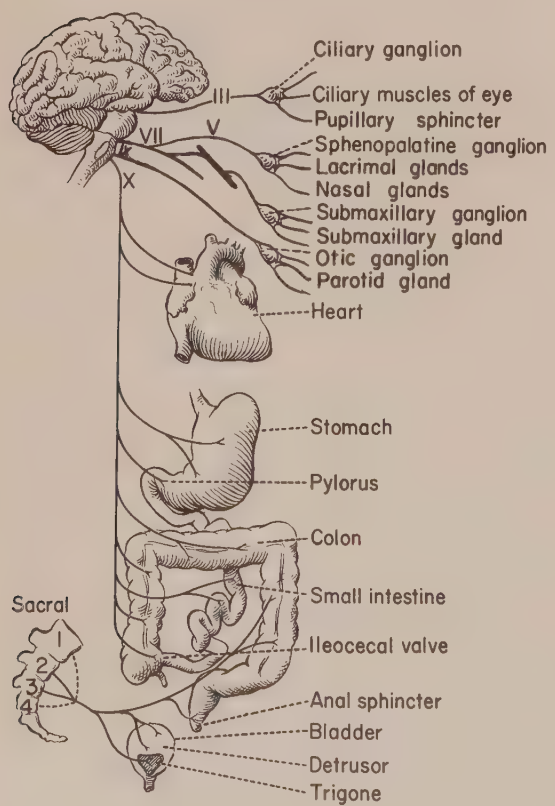


Figure 255. Anatomy of the parasympathetic nervous system.

Adrenergic versus cholinergic nerves in the autonomic nervous system. One of the major differences between parasympathetic and sympathetic nerves is that the postganglionic neurons of the two systems usually secrete different hormones. The postganglionic neurons of the parasympathetic system secrete *acetylcholine,* for which reason these neurons are said to be *cholinergic.* Those of the sympathetic system secrete mainly norepinephrine; another name for norepinephrine is noradrenaline, for which reason most postganglionic neurons of the sympathetic system are said to be *adrenergic.* However, a few are cholinergic like the parasympathetic neurons, as is noted below.

Actions of the Parasympathetic and Sympathetic Nervous Systems on Different Organs

Effect on the eye. Table 6 lists the functions of the parasympathetic and sympathetic systems with respect to different organs. This table shows that the sympathetic system dilates the pupil of the eye, allowing increased quantities of light to enter, and the parasympathetics constrict the pupil, thus decreasing the amount of light. The parasympathetics also control the ciliary muscle that focuses the lens for far and near vision.

Secretion of the digestive juices. The secretion of digestive juices by some of the glands of the gastrointestinal tract is controlled by the parasympathetics, while the sympathetics have very little effect, if any at all. The salivary glands of the mouth and the fundic glands of the stomach normally are almost entirely controlled by the parasympathetics. On the other hand, the glands in the intestines are controlled to a slight extent by the parasympathetics, but mainly by local effects in the intestines themselves.

Effects on the sweat glands. The sweat glands are stimulated by fibers from the sympathetic nervous system. However, these fibers are different from the usual sympathetic fibers, for they

Table 6. *Autonomic Effects on Various Organs of the Body*

Organ	*Effect of Sympathetic Stimulation*	*Effect of Parasympathetic Stimulation*
Eye: Pupil	Dilated	Contracted
Ciliary muscle	None	Excited
Gastrointestinal glands	Vasoconstriction	Stimulation of thin, copious secretion containing many enzymes
Sweat glands	Copious sweating (cholinergic)	None
Heart: Muscle	Increased activity	Decreased activity
Coronaries	Vasodilated	Constricted
Systemic blood vessels:		
Abdominal	Constricted	None
Muscle	Dilated (cholinergic)	None
Skin	Constricted or dilated (cholinergic)	None
Lungs: Bronchi	Dilated	Constricted
Blood vessels	Mildly constricted	None
Gut: Lumen	Decreased peristalsis and tone	Increased peristalsis and tone
Sphincters	Increased tone	Decreased tone
Liver	Glucose released	None
Kidney	Decreased output	None
Bladder: Body	Inhibited	Excited
Sphincter	Excited	Inhibited
Male sexual act	Ejaculation	Erection
Blood glucose	Increased	None
Basal metabolism	Increased up to 150%	None
Adrenal cortical secretion	Increased	None
Mental activity	Increased	None

are mainly cholinergic rather than adrenergic. Also, they are stimulated by the nervous centers in the brain that control the parasympathetics, rather than by the centers that control the sympathetics. Therefore, despite the fact that the fibers supplying the sweat glands are anatomically sympathetic, they can be considered physiologically to be parasympathetic fibers.

Effects on the heart. Stimulation of the sympathetic nervous system increases the heart's activity and also dilates the coronaries so that increased nutrition will be available. On the other hand, parasympathetic stimulation decreases the activity of the heart while constricting the coronaries to conserve blood flow. These effects were discussed in detail in Chapter 12.

Control of the blood vessels. Perhaps the most important function of the sympathetic nervous system is to control the blood vessels in the body. Most vessels are constricted by sympathetic stimulation, though a few, the coronaries and skeletal muscle vessels for instance, are dilated. By controlling the peripheral blood vessels, the sympathetic nervous system is capable of regulating both the cardiac output and arterial pressure; constriction of the veins and venous reservoirs increases the cardiac output, and constriction of the arterioles increases the peripheral resistance, which elevates the arterial pressure.

The parasympathetics, when they affect the blood vessels at all, dilate them, but this effect is so slight and occurs in such a few areas of the body, the brain for instance, that it can be almost totally ignored.

Effect on the lungs. The bronchi are dilated by sympathetic stimulation or by administration of a sympathetic-like drug. However, neither the parasympathetic nor sympathetic system has any significant effect on the blood vessels of the lungs. This is different from the effects on the blood vessels of the remainder of the body.

Control of gastrointestinal movements.

About 75 per cent of all the parasympathetic nerve fibers are distributed to the gastrointestinal tract, which indicates that by far the most important function of this entire system is regulation of gastrointestinal activities. Parasympathetic stimulation increases peristalsis and at the same time decreases the tone of the sphincters. Peristalsis propels the food forward while the open sphincters between the different segments of the gastrointestinal tract allow the food to move forward with ease. During extreme parasympathetic stimulation, food can actually pass all the way from the mouth to the anus in approximately 30 minutes, though the normal transmission time is about 24 hours.

Sympathetic stimulation, on the other hand, inhibits peristalsis and tightens the sphincters. This slows down the movement of food through the gastrointestinal tract.

Release of glucose from the liver. Sympathetic stimulation causes rapid breakdown of glycogen into glucose in the liver and then release of the glucose into the blood. This increased level of glucose in the blood provides a quick supply of nutrition for the tissue cells, an effect especially valuable during exercise.

Effect on the kidneys. Sympathetic stimulation causes intense vasoconstriction of the renal blood vessels and greatly decreases the output of urine. This is a very important mechanism for the regulation of blood volume and arterial pressure, for when need be, sympathetic stimulation can cause fluid to be retained in the circulatory system, increasing the blood volume and the venous return to the heart. These effects in turn increase the cardiac output and raise the blood pressure to the desired level.

Emptying of the bladder. Emptying of the bladder is caused mainly by the parasympathetic system, which excites the muscular wall of the bladder while inhibiting the sphincters that normally prevent the flow of urine. On the other hand, sympathetic stimulation prevents emptying of the bladder. Though the sympathetic effect ordinarily is not important, occasionally when a person has severe peritoneal inflammation in the region of the bladder, a resulting peritoneal reflex excites the sympathetics so greatly that the person becomes unable to urinate.

Control of sexual functions. The autonomic nervous system also enters into the control of the sexual acts of both the male and the female. In the male the parasympathetics cause erection, and the sympathetics cause ejaculation. In the female the parasympathetics cause erection of the erectile tissue around the vaginal opening, which causes tightening, and they also cause the female to secrete large quantities of mucus which facilitates the sexual act. The effect of the sympathetics on the female sexual act is not well understood, but it is believed that these nerves initiate reverse uterine peristalsis during the female climax.

Metabolic effects. Generalized sympathetic stimulation increases the metabolism of all cells of the body. The sympathetic nerve fibers are so widespread in all tissues that at least some norepinephrine secreted by their nerve endings seems to reach every functioning cell. The norepinephrine increases the rates of the chemical reactions in all cells, thereby increasing the overall rate of metabolism of the body. In this way the sympathetics can keep a person warm when he tends to become too cold, and during exercise or other states of activity they can make the body perform greater quantities of work than would be possible otherwise. These metabolic effects can be brought about in only a few seconds, and they can be stopped in another few seconds when the need for increased metabolism is over.

Stimulation of the adrenal cortex. When the body is subjected to physical stress of sufficient intensity to damage tissues, the sympathetics usually become stimulated, and this initiates the secretion of adrenocortical hormones in the following manner: Stimulation of the sympathetic nerve endings in the anterior pituitary gland causes this gland to secrete a hormone called *adrenocorticotropic hormone,* which in turn flows by way of the blood stream to the adrenal cortex where it promotes the output of adrenocortical hormones. These hormones then help the tissue cells repair damage that has resulted from the stress.

Stimulation of mental activities. A final important effect of sympathetic stimulation is an increase in the rate of mental activity. This probably results from the increased rate of metabolism in the neuronal cells.

Overall Function of the Sympathetic Nervous System

In general, when the sympathetic centers of the brain become excited they stimulate almost all of the sympathetic nerves at once. As a result, the blood pressure rises, the rate of metabolism increases, the degree of mental activity is enhanced, and the glucose in the blood increases. Considering all of these effects together, it can be seen that mass discharge of the sympathetic nervous system

prepares the body for activity. When a person's sympathetic nerves have been destroyed he cannot "generate steam"; in other words, he simply cannot reach the high levels of vigor often necessary to perform rapid, forceful, and excited activities.

Special functions of the sympathetic nervous system. In addition to the mass discharge mechanism of the sympathetic system, certain physiological conditions can stimulate localized portions of the system independently of the remainder. Two of these are the following:

HEAT REFLEXES. Heat applied to the skin causes a reflex through the spinal cord and then back again to the skin to dilate the blood vessels in the heated area. Also, heating the blood flowing through the heat control centers of the hypothalamus increases the degree of vasodilatation of all the skin blood vessels without significantly affecting the vessels deep in the tissues. Most of these reactions are effected through vasodilator sympathetic fibers to the skin.

SHIFT OF BLOOD FLOW TO THE MUSCLES DURING EXERCISE. During exercise two separate portions of the sympathetic nervous system are stimulated, the vasodilator fibers to the muscles and the vasoconstrictor fibers to most other parts of the body. Nerve impulses pass directly from the motor cortex to the sympathetic regions of the brain stem to initiate these effects. Vasodilatation in the muscles allows blood to flow through them with great ease, while vasoconstriction elsewhere, except in the heart and brain, decreases almost all other bodily blood flow. Thus, in exercise the sympathetic nervous system causes a massive shift of blood flow to the active muscles.

Function of the adrenal medulla. The adrenal medulla is the central portion of the adrenal gland. The medulla is surrounded by the adrenal cortex, which was mentioned above. These two portions of the adrenal gland are actually two separate glands, for they secrete entirely different types of hormones, and regulation of each is entirely different from the other.

The glandular cells of the medulla are modified postganglionic sympathetic neurons that secrete norepinephrine and epinephrine directly into the blood when stimulated by sympathetic impulses. These hormones then flow throughout the body, causing sympathetic effects everywhere. Therefore, every time a mass discharge occurs in the sympathetic nervous system, in addition to direct stimulation of the organs by the sympathetic nerves, large quantities of norepinephrine and epinephrine also flow to all points in the body. These hormones then exert exactly the same effects on the different organs as direct sympathetic stimulation by the nerves. For instance, they increase the activity of the heart, inhibit peristalsis in the gut, increase the rate of metabolism of all cells, and so forth. In effect, then, the adrenal medullae provide a second means for causing all the sympathetic activities. They also allow those parts of the body that do not have sympathetic nerve fibers to respond to sympathetic stimulation.

The quantity of norepinephrine and epinephrine secreted by the adrenal medullae is great enough so that about one half of all the functions of the sympathetic nervous system are mediated by this means rather than by direct stimulation of the different organs. In fact, when all the sympathetic nerves except those to the adrenal medullae are destroyed, the sympathetic system will still function almost normally because the adrenals compensate by secreting extra quantities of hormones. Conversely, loss of adrenal medullary secretion is hardly discernible because of compensation by the nerves. To stop entirely the action of the sympathetic nervous system, the function of both the adrenal medullae and all the nerve fibers that go directly to the organs must be blocked.

Parasympathetic and Sympathetic Tone

Impulses normally are transmitted continuously, though at a relatively slow rate, through all the fibers of both the parasympathetic and sympathetic systems. This allows continuous stimulation of the internal structures, which is called *tone* or the *tonic effect* of the system. Tone allows each system to exert both positive and negative effects on a structure—that is, by increasing the number of impulses above the normal value, the effect can be increased; on the other hand, by decreasing the number of impulses below the normal value the effect can be decreased. For instance, if the sympathetic vasoconstrictor fibers to the blood vessels were normally dormant, it would be possible for sympathetic regulation only to constrict the vessels and never to dilate them; however, because of the normally persistent tone, the blood vessels are always partially constricted so that the sympathetics can either further constrict the vessels by increasing their stimulation or they can dilate the vessels by decreasing their stimulation. These principles apply throughout the parasympathetic and sympathetic systems and are responsible for a higher degree of effectiveness than would be possible otherwise.

CONTROL OF THE AUTONOMIC NERVOUS SYSTEM BY NERVE CENTERS IN THE BRAIN

Not much is known about the exact centers in the brain that regulate the discrete autonomic functions. However, stimulation of widespread areas in the reticular substance of the medulla, pons, and mesencephalon can cause all of the effects of the sympathetic and parasympathetic systems. The hypothalamus can regulate many, though not all, of the functions of the autonomic system. Finally, stimulation of localized areas of the cerebral cortex can cause general or sometimes discrete autonomic effects.

Important autonomic centers of the brain stem. Figure 256 shows a tentative outline of the centers in the hindbrain and hypothalamus that regulate different autonomic functions of the internal organs. It will be noted from this figure that the vascular system and respiration are controlled mainly by areas in the reticular substance of the medulla and pons. The urinary bladder is controlled by centers in the mesencephalon, but these centers are usually suppressed by areas in the cerebral cortex that provide conscious regulation of bladder emptying.

Autonomic centers of the hypothalamus. The centers of the hypothalamus are concerned mainly with two features of autonomic regulation. These are, first, regulation of metabolism, and, second, regulation of body heat. Stimulation of a small area in the anterior hypothalamus increases a person's appetite, while stimulation of the posterior hypothalamus usually greatly inhibits his appetite. In this way the diet and its quality are regulated. Also, stimulation of a small region in the anterior hypothalamus helps to regulate water metabolism by causing the secretion of antidiuretic hormone. This hormone in turn makes the kidneys reabsorb large quantities of water.

Regulation of body heat by the hypothalamus is vested in opposing parasympathetic and sympathetic centers. Stimulation of the anterior hypothalamus excites the parasympathetic system, while stimulation of the posterior hypothalamus excites the sympathetic system. When the blood temperature falls too low, the posterior hypothalamus becomes excited, which in turn initiates sympathetic effects throughout the body. These include (1) an increase in the rate of metabolism, which increases the liberation of heat, and (2) increased vasoconstriction of all the blood vessels of the skin to prevent loss of heat from the body. As a result, the body temperature rises to normal. On the other hand, when the blood temperature rises too high, the posterior hypothalamus becomes inhibited while the anterior hypothalamus becomes stimulated. The resultant impulses induce sweating throughout the body, causing a refrigeration effect, and stimulation of the vasodilator fibers (cholinergic) to the skin causes the blood vessels in the skin to dilate so that large quantities of heat can be lost. These effects decrease the body temperature back toward normal. Heat control will be considered in greater detail in Chapter 32.

Important cerebral centers. Stimulation of the autonomic nervous system by the cerebral

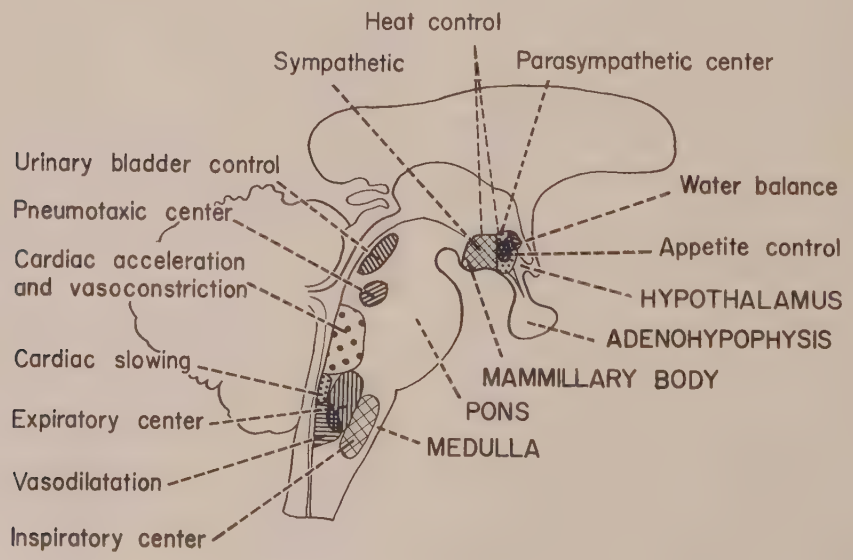


Figure 256. The major autonomic centers in the brain stem.



Figure 257. The reticular activating system, showing by the arrows passage of impulses from the reticular substance of the mesencephalon upward through the thalamus to all parts of the cortex.

cortex occurs principally during emotional states. Many discrete centers, especially in the prefrontal lobes and temporal regions of the cortex, can increase or decrease the degree of excitation of the hypothalamus centers. Also, the thalamus and closely related structures deep in the cerebrum help to regulate the hypothalamus. Therefore, both conscious and subconscious portions of the cerebrum can cause autonomic effects. Sudden psychic shock, which originates in the cerebral cortex, can cause fainting. This usually results from powerful stimulation of the vasodilator fibers and the cardiac inhibitory fibers (vagus nerves) which cause the arterial pressure to fall precipitously. On the other hand, extreme degrees of psychic excitement, which can result from either conscious or subconscious stimulation in the cerebrum, can stimulate the vasomotor system to increase the blood pressure and cardiac output.

Unfortunately, our present knowledge of the relation of the cerebrum to the autonomic nervous system is so scanty that it can be only descriptive rather than explanatory of the actual functional relationships.

SLEEP AND WAKEFULNESS

One of the most important mysteries of brain function is its diurnal cycle of sleep and wakefulness. Even when a person remains in total darkness or in total light he still maintains approximately the same sleep-wakefulness cycle, with a periodicity of once every 24 hours. Ordinarily the nervous system shows signs of fatigue shortly before sleep ensues, and it shows signs of having become considerably rested at the termination of sleep. It seems, then, that neuronal fatigue plays an important part in causing sleep, and that sleep in turn relieves the fatigue.

Electrical studies of the brain indicate that while a person is awake many nerve impulses pass continuously through the nervous system, never ceasing. However, when the person is asleep only a few straggling impulses are present at any one time. Thus, the state of wakefulness seems to be caused by a high degree of activity in the cerebrum, while the state of sleep is caused by a low degree of activity. Therefore, any theory to explain sleep and wakefulness must also explain these changing degrees of cerebral activity during the two states.

The reticular activating system. Stimulation of portions of the brain stem and thalamus greatly increases the activity of the cortex, an effect shown diagrammatically in Figure 257. Stimulation anywhere along the arrows will cause impulses to spread upward and eventually to excite the cortex. This system is called the *reticular activating system,* and it is divided into two separate parts, the *mesencephalic part* and the *thalamic part.*

The mesencephalic part of the reticular activating system is composed mainly of the reticular substance in the mesencephalon and the reticular substance of the upper pons. Stimulation of this region causes very diffuse flow of impulses upward through widespread areas of the thalamus and thence to widespread areas of the cortex, causing generalized increase in cerebral activity.

The thalamic portion of the reticular activating

system differs from the mesencephalic system in that stimulation here activates localized regions of the cerebral cortex. Stimulation posteriorly in the thalamus activates posterior parts of the cerebral cortex while stimulation anteriorly activates anterior parts of the cortex. Thus, signals from specific parts of the thalamus can call forth activity in specific parts of the cortex rather than activating the entire cortex.

It is likely that the function of the thalamic portion of the reticular activating system is to direct one's attention to memories or thoughts stored in specific parts of the cerebral cortex and thereby to allow his consciousness to consider individual thoughts at a time. On the other hand, activation of the mesencephalic portion of the reticular activating system causes generalized wakefulness.

THE AROUSAL REACTION. The reticular activating system itself must be stimulated to action by input signals from other sources. When an animal is asleep, the reticular activating system is in an almost totally dormant state, yet almost any type of sensory signal can immediately activate the system. For instance, pain stimuli from any part of the body, or proprioceptor signals from the vestibular apparatuses, from the joints, and so forth can all cause immediate activation of the reticular activating system. This is called the *arousal reaction,* and it is the means by which sensory stimuli awaken us from deep sleep.

CEREBRAL STIMULATION OF THE RETICULAR ACTIVATING SYSTEM. Signals from the cerebral cortex can also stimulate the reticular activating system and thereby increase its activity. Fiber pathways to both the mesencephalic and thalamic portions of this system are particularly abundant from the *somesthetic cortex,* the *motor cortex,* the *frontal cortex,* and parts of the cortex that deal mainly with emotions, the *cingulate gyrus* and the *hippocampus.* Whenever any of these cortical regions becomes excited, impulses are transmitted into the reticular activating system, thereby increasing the degree of activity of the reticular activating system.

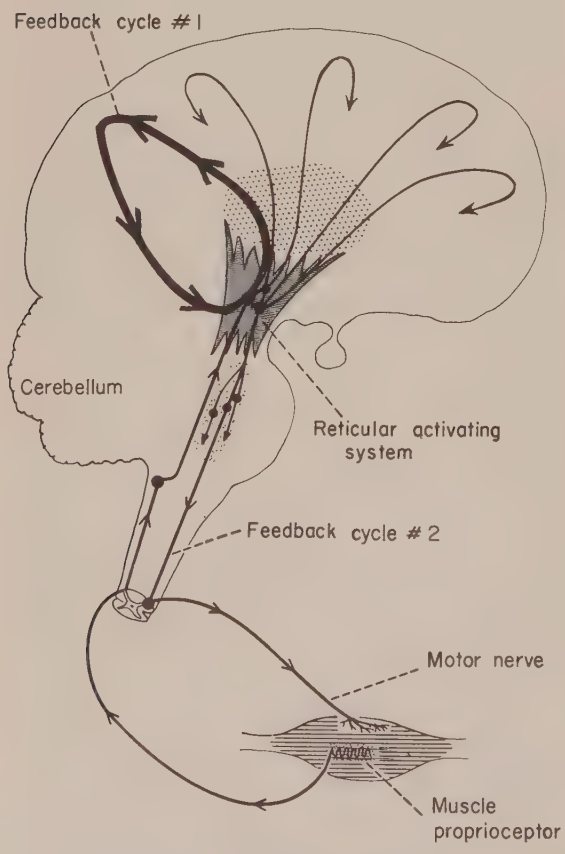


Figure 258. The feedback theory of wakefulness, showing two feedback cycles passing through the reticular activating system: one to the cerebrum and back (#1), and another to the peripheral muscles and back (#2).

The Feedback Theory of Wakefulness and Sleep

From the above discussion we can see that activation of the reticular activating system intensifies the degree of activity of the cerebral cortex, and in turn increased activity in the cerebral cortex increases the degree of activity of the reticular activating system. Therefore, it is fairly evident that a so-called "feedback loop" could develop whereby the reticular activating system excites the cortex and the cortex excites the reticular activating system, thus setting up a cycle that causes continued intense excitation of both of these regions. Such a cycle is shown in Figure 258.

The reticular activating system is also involved in another feedback loop as follows: It sends impulses down the spinal cord to increase the tone in all the muscles of the body. In turn, the increased tone excites muscle proprioceptors that send sensory impulses upward to excite the reticular activating system. Thus, here again, activity of the reticular activating system sets off another cycle that in turn further increases the activity in the reticular activating system. This cycle is also shown in Figure 258.

One can readily see, therefore, that once either or both of these feedback loops involving the reticular activating system should become excited, the increased activity in the reticular activating system would stimulate all other parts of the brain, thereby creating a state of wakefulness.

On the other hand, one might wonder how *sleep* could possibly occur once the different feedback loops should become activated. A probable answer to this is that all synapses of the central nervous system eventually fatigue, so that they either stop or nearly stop transmitting impulses. Therefore, after a prolonged period of wakefulness, one would expect the synapses in the feedback loops to become fatigued and therefore the feedback loops to stop functioning, thus allowing the reticular activating system to become dormant, which is the state of *sleep.*

We also know that a person can exist in all levels of wakefulness and sleep. The feedback theory of wakefulness postulates that the degree of wakefulness depends on the number of feedback loops activated at any one time. Figure 258 illustrates a large number of activating pathways passing from the mesencephalic activating system upward through the thalamus to the cortex. If only one of these pathways becomes activated, then obviously the degree of wakefulness would be very slight. If large numbers of the pathways should become activated simultaneously, then the degree of wakefulness would be greatly enhanced. One would expect that after a long period of wakefulness, progressively more and more of the feedback loops would become inactive and a person would fade into a state of sleep. Then upon arousal all of the feedback loops might be activated simultaneously, which explains why one could go from a state of sleep to complete wakefulness in a few seconds.

Effects of Wakefulness

Wakefulness is associated with three major effects: (1) increase in the degree of activity of the cerebrum, (2) transmission of signals directly from the reticular activating system to the muscles, and (3) excitation of the sympathetic nervous system.

The increase in cerebral function is caused by the great number of impulses transmitted from the wakefulness center upward through the thalamus to the cerebral cortex. These impulses continually impinge on the cerebral neurons, facilitating them to activity.

Stimulation of the reticular areas of the brain stem increases the degree of tone in all the muscles of the body. This makes the muscles more excitable than would otherwise be true and prepares them for immediate activity, which explains why a person who is wide awake has the feeling of muscular readiness.

The sympathetic stimulation caused by wakefulness increases the blood pressure slightly; it increases the rate of metabolism in all the tissues of the body; and, in general, it simply makes the body ready to perform increased amounts of work.

Effects of Sleep

Sometimes it is hard to understand why a person needs to sleep at all. Certain parts of the body such as the heart never rest and still are capable of functioning throughout life. One might reason that sleep is a measure to conserve the energies of most parts of the body when they are not needed. Some animals, as a matter of fact, have carried this principle of conservation so far that they pass into a state of very prolonged and often deep sleep called *hibernation,* which lasts throughout the entire winter.

A special value of sleep seems to be to reestablish appropriate balance of excitability among the various portions of the nervous system. As a person becomes progressively fatigued, some parts of his central nervous system lose excitability more than others, so that one part may overbalance the others. In fact, extreme fatigue can even precipitate severe psychotic disturbances. Yet after prolonged sleep, all parts of the nervous system usually will have returned once again to appropriate degrees of excitability and to a state of serenity.

Lack of sleep does not *directly* affect the intrinsic functions of the different organs. However, lack of sleep often causes severe autonomic disturbances, and these in turn *indirectly* lead to gastrointestinal upsets, loss of appetite, and other detrimental effects. In this way loss of sleep can affect the whole body as well as the nervous system itself.

Brain Waves

Electrical impulses called *brain waves* can be recorded from all active parts of the brain and even from the outside surface of the head. The character of these waves is closely related to the degree of sleep and wakefulness. When a person is awake but not thinking hard, continuous waves at a rate of approximately 10 to 12 per second can be recorded from almost all parts of the cerebral cortex. These are called *alpha waves.* The impulses that cause them probably originate in the reticular activating system and then spread into the cerebral cortex. These are believed to be the signals that keep the cortex facilitated during wakefulness. The alpha waves are illustrated by the recording at the top of Figure 259.

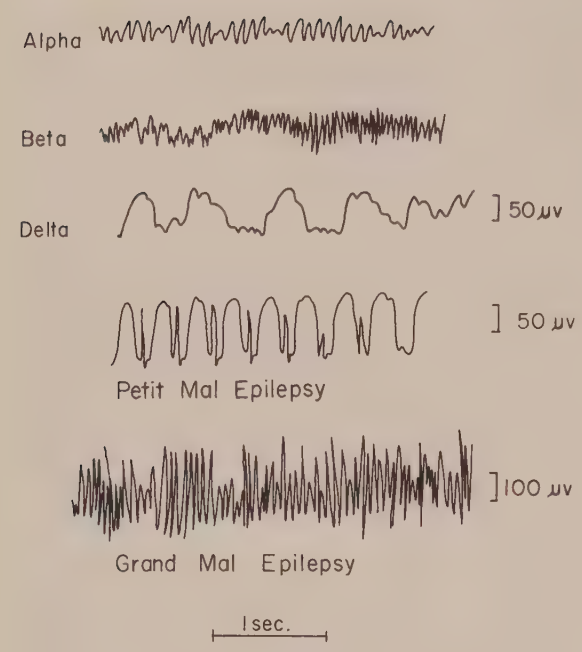


Figure 259. Brain wave patterns in the normal person and in two persons with different types of epilepsy.

When any part of the brain becomes very active, for example the motor region initiating muscular activities, additional waves having a frequency sometimes as high as 50 cycles per second, and intensities usually much greater than those of the alpha waves, take the place of the normal alpha waves. These, called *beta waves,* are shown by the second recording of Figure 259.

During sleep the alpha and beta waves cannot be recorded and in their place a few straggling waves occur approximately once every one to two seconds. These are the "sleep waves" or *delta waves,* as shown by the third recording of Figure 259.

Abnormal brain wave patterns. Various abnormalities of the brain can cause strange brain wave patterns. Two of these, caused by different types of *epilepsy,* are shown in the fourth and fifth records of Figure 259. The fourth record shows a "spike and dome" picture which occurs in *petit mal* epilepsy. In this disease the person suddenly becomes unconscious for 3 to 10 seconds at a time. Such episodes may occur every few minutes, every few hours, or only once in many months, and when they do occur the person usually continues, while unconscious, whatever physical activity he is already doing even though he might be walking across a crowded street. Petit mal epilepsy seems to result from some abnormality of the wakefulness regions of the brain. The transmission of the normal alpha waves to the cerebral cortex is temporarily stopped. Instead, the spike and dome pattern is transmitted, and the person falls asleep for a few seconds until the alpha wave pattern picks up again.

The bottom record of Figure 259 shows the brain waves in *grand mal* epilepsy. In this condition the cerebral cortex becomes extremely excited, and many very strong signals spread over the brain at the same time. When these reach the motor cortex they cause rhythmic movements, called *clonic convulsions,* throughout the body. Grand mal epilepsy probably results from abnormal reverberating cycles developing in the reticular activating system. That is, one portion of the system stimulates another portion, this stimulates a third portion, and this in turn restimulates the first portion, causing a cycle that continues for two to three minutes until the neurons of the system fatigue so greatly that the reverberation ceases. At the beginning of a grand mal attack a person may experience very violent, abnormal, hallucinatory thoughts; once most of the brain is involved, he no longer has any conscious thoughts at all, because signals are then being transmitted in all directions rather than through discrete thought circuits. Therefore, even though his brain is violently active, he becomes unconscious of his surroundings. Following the attack his brain is so fatigued that he sleeps at least a few minutes and sometimes as long as several days.

THE EMOTIONS

The emotions are one's inner "feelings" that can be greatly affected by mental activities over which the person has no voluntary control. The involvement of different portions of the brain in the emotions may be described as follows:

Part played by the cerebrum. Since it is the cerebral cortex that is mainly responsible for thought, it is here also that the centers for the emotions are located. During operations on the human brain it has been possible to stimulate various cerebral areas of patients while they are awake, and to determine the resultant sensations. Also, many similar experiments have been performed in monkeys, from which at least some indications of emotions can be discerned. Stimulation of different prefrontal regions of the brain results in varying emotions such as depression, worry, excitement, and anxiety. On the other hand, stimulation of the cingulate gyrus, located deep in the longitudinal fissure, can cause docility and fear. And, finally, stimulation of areas in the temporal lobes often causes the emotion of rage.

Another part of the cerebrum, the *amygdaloid nuclei,* which lie laterally inside the two hemispheres and beneath the cortex, is also important in control of behavioral and emotional patterns. Stimulation of different parts of the amygdaloid nuclei can cause such emotions and behavior as fear, alert attention, fighting activities, vomiting, and sexual activities. Thus, it is believed that the normal function of the amygdaloid nuclei is to help control the overall pattern of behavior demanded for each occasion. The amygdaloid nuclei cause different behavioral activities mainly by stimulating different parts of the hypothalamus.

Part played by the thalamus. The medial nuclei of the thalamus seem to be very important in determining whether one's thoughts are pleasant or unpleasant. For this reason, the thalamus probably helps to set the general tenor of the emotions. When the thalamic nuclei are damaged, very intense emotions are often experienced, sometimes anger and at other times extreme happiness.

After the cerebral cortex has been removed from an animal, the animal can still exhibit signs of emotion. These probably result mainly from thalamic and hypothalamic function. Such an animal can be shown food and will exhibit evidences of desiring the food by licking his lips and beginning to salivate. He can also be stimulated, by certain auditory or somesthetic sensations, to exhibit a state of rage.

Relationship of the hypothalamus to the emotions. The hypothalamus operates at a subconscious level, but it can cause or modify emotions by transmitting signals to the thalamus and then to the cerebral cortex. For instance, intense stimulation of the posterior hypothalamus excites greatly the entire cerebral cortex, and, therefore, can cause extreme excitement. On the other hand, destruction of the posterior hypothalamus can cause a state of almost total lethargy. Also, stimulation of other regions of the hypothalamus can affect the appetite either positively or negatively, depending on the precise point stimulated.

Reward and Punishment

One of the most important recent discoveries in the field of behavior is the so-called "reward and punishment" system of the brain. Certain areas in the thalamus, hypothalamus, and mesencephalon, when stimulated, make the animal feel intense punishment as if he were being severely pained. Yet stimulation of other closely related areas cause exactly the opposite effect, making the animal appear to be experiencing extreme pleasure. One of the experimental methods for studying the reward and punishment centers is to implant an electrode in one or the other of these two areas and then to allow the animal itself to press a lever to control the stimulus. If the electrode is implanted in the reward area, the animal will press the lever continually. Indeed, it would rather stimulate its reward center than eat, even though it might be starving. On the other hand, if the electrode is placed in the punishment area, it will avoid stimulation by all means possible.

Stimulation of the punishment region usually makes the animal become extremely angry and often causes a state of severe rage. On the other hand, stimulation of the reward areas causes the animal to become completely docile and to be amenable to almost any type of treatment.

Relationship of reward and punishment to learning. Experiments on learning and memory have shown that an animal remembers sensory stimuli that cause either reward or punishment but fails to remember sensory stimuli that fail to excite either the reward or punishment area. For instance, a food that is very pleasant to the taste is remembered, or, likewise, a food that is exceedingly unpleasant is remembered. On the other hand, food that causes neither pleasure nor displeasure is rapidly forgotten. Similarly, a very painful stimulus, such as touching a hot iron, is remembered well, whereas simply touching a book or stick of wood is forgotten in a few seconds.

Therefore, in the process of memory, two different components must be present for a sensory experience to be remembered. The first component is the sensory experience itself, and the second component is an experience of either reward or punishment—that is, an experience of pleasure or displeasure.

PSYCHOSOMATIC EFFECTS

A psychosomatic effect is a bodily, or *somatic,* effect produced by psychologic stimulation. The brain can produce such effects in three general ways: (1) by transmission of signals through the autonomic nervous system, (2) by transmission of signals to the muscles through the bulboreticular area, and (3) by control of certain of the endocrine glands. Figure 260 illustrates the general neuronal mechanisms believed to be responsible for psychosomatic effects. It shows conscious signals beginning either in the frontal cortex or in the thalamus, then being transmitted to the hypothalamus, and finally going through the centers of the hindbrain to the somatic regions.

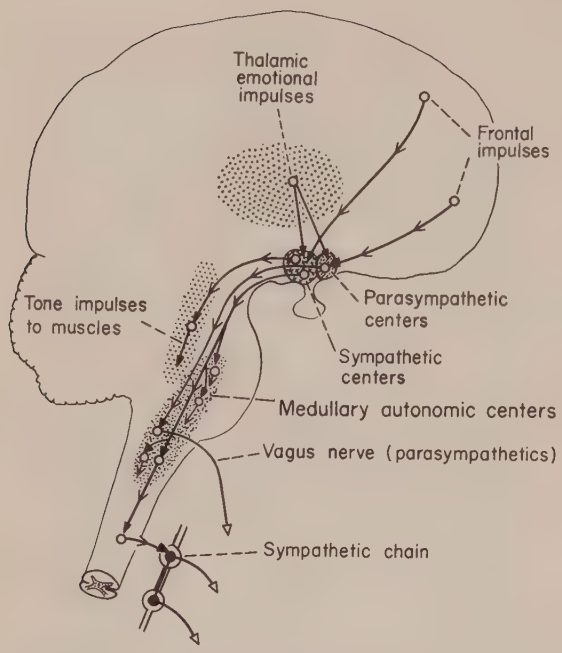


Figure 260. Neuronal circuits responsible for psychosomatic effects.

Psychosomatic effects transmitted through the autonomic nervous system. Almost any emotion can affect the autonomic nervous system. For example, a very intense emotion will often increase the excitability of the reticular activating system, and because this system also controls the sympathetics, it excites sympathetic activity throughout the body. Therefore, generalized sympathetic stimulation of the organs is one of the most common of all psychosomatic effects. The emotions of excitement, anxiety, rage, and so forth often discharge the sympathetics *en masse,* causing marked increase in arterial pressure, palpitation of the heart, and cold chills over the skin.

Psychologic effects also often stimulate the parasympathetic centers of the hypothalamus. The emotions of worry, depression, and lethargy, all of which have effects opposite to those that excite the sympathetic system, usually stimulate the parasympathetics. On occasion, however, both of the systems may be stimulated simultaneously. Fear, for instance, can cause extreme sympathetic stimulation resulting in elevation of arterial pressure, while at the same time stimulating the parasympathetics to cause such intense gastrointestinal activity that the person has diarrhea.

Transmission of psychosomatic effects through the skeletal muscles. The reticular activating system sends signals downward through the cord and directly to the muscles. Therefore, the same emotions that excite the sympathetics and reticular activating system usually increase the tone of the muscles throughout the body. Sometimes the tone becomes so intense that it causes muscle tremor, which explains why certain emotions can culminate in actual shaking.

On the other hand, the emotions that normally stimulate the parasympathetics usually decrease the activity of the bulboreticular formation. As a result, the muscular tone decreases to a very low level, which explains the muscular *asthenia* (muscular weakness) that is characteristic of some psychic states.

Transmission of psychosomatic effects through glands. The nervous system controls several of the endocrine glands either completely or partially. For instance, the sympathetic nervous system controls the adrenal medulla, and the hypothalamus controls almost all the activities of the adenohypophysis. The hypothalamus effects its control by secreting *neurosecretory substances* within the substance of the hypothalamus; these are carried through minute veins from the hypothalamus to the adenohypophysis where they cause secretion of several major hormones, *growth hormone, corticotropin, thyrotropin,* and three *gonadotropins.* These hormones in turn control growth rate, protein metabolism, overall rate of metabolism, and most sexual functions. Also, the hypothalamus controls the secretion of *antidiuretic hormone* by neurons located in the supraoptic nuclei; this hormone in turn controls the degree of retention of water by the kidneys.

Obviously, therefore, many psychosomatic effects can be mediated through the endocrine glands. For instance, psychic effects that overly stimulate the hypothalamus can cause hyperthyroidism, the thyroid gland secreting excess thyroid hormone and thereby increasing the rate of metabolism of all cells of the body. Likewise, psychic signals can affect the output of sex hormones and thereby cause failure of ovulation, excess menses, diminished menses, infertility, and other sexual abnormalities.

Psychosomatic Diseases

Occasionally abnormal cerebral function can cause severe physical disorders. Perhaps the most common psychosomatic disease is extreme tension throughout the body, resulting from excessive stimulation of the sympathetics and from increased tone of the peripheral muscles. This is often described simply as a "nervous breakdown."

Psychosomatic effects can also cause abnormal function of individual organs. For instance, stimulation of the sympathetics can so decrease gas-

trointestinal activity that constipation results. On the other hand, excessive stimulation of the parasympathetics can so increase the degree of gastrointestinal activity as to cause severe diarrhea. A very common psychosomatic disorder is palpitation of the heart caused by excitement, anxiety, or other emotional states.

Occasionally the dysfunction caused by a psychosomatic disorder is so great that tissues are actually destroyed. For instance, stimulation of the parasympathetics to the stomach can cause so much secretion of gastric juices that they eat a hole into the wall of the stomach or upper intestine. This causes the condition called *peptic ulcer.* Such patients can be treated by operative removal of portions of the stomach, by neutralizing the gastric juices with special drugs, or in some instances by psychiatric treatment to alleviate the emotional condition that is initiating the excessive secretion of gastric juices.

Psychosomatic pain. Many psychosomatic disorders can lead to pain which is called simply psychosomatic pain. For instance, a stomach ulcer causes intense burning in the pit of the stomach. Spasm of the gut caused by excess parasympathetic stimulation causes cramps in the abdomen. And, occasionally, overstimulation of the heart can even cause cardiac pain. If the psychic condition that causes the functional abnormality can be corrected, then the pain likewise will be corrected.

Myths about psychosomatic disease. Despite the many different ways in which psychosomatic disease can come about, this subject has been greatly overemphasized by persons not familiar with it. It is a common myth that a person can worry so much about the function of one of his organs that he thereby creates disease in that particular organ. Except in a few instances this is not true. Most psychosomatic diseases exhibit regular patterns such as general states of tension, constipation or diarrhea, ulcer, and a few others of a similar nature.

REFERENCES

Axelrod, J.: Metabolism of epinephrine and other sympathomimetic amines. *Physiol. Rev.* 39:751, 1959.

Brooks, F. P.: Visceral functions of the nervous system. *Ann. Rev. Physiol.* 22:473, 1960.

Burnstock, G., and Holman, M. W.: Smooth muscle: autonomic nerve transmission. *Ann. Rev. Physiol.* 25:61, 1963.

Ciba Foundation, Symposium Volume: Neurological Basis of Behavior. Boston, Little, Brown and Company, 1958.

Eccles, J. C.: *Neurophysiological Basis of Mind: The Principles of Neurophysiology.* London, Oxford University Press, 1953.

Gellhorn, E.: *Physiological Foundations of Neurology and Psychiatry.* Minneapolis, The University of Minnesota Press, 1953.

Olds, J.: Higher functions of the nervous system. *Ann. Rev. Physiol.* 21:381, 1959.

Pennes, H. H.: *Progress in Neurobiology.* Vol. 3. *Psychopharmacology: Pharmacologic Effects on Behavior.* New York, Paul B. Hoeber, Inc., 1958.

Poynter, F. N. L.: *The History and Philosophy of Knowledge of the Brain and Its Functions.* Springfield, Ill., Charles C Thomas, 1959.

Pribram, K. H.: The intrinsic systems of the forebrain. *Handbook of Physiology* 1:1323, 1960.

Strom, G.: Visceral functions of the nervous system. *Ann. Rev. Physiol.* 23:419, 1961.

Section Seven

THE SPECIAL SENSORY SYSTEMS

THE EYE

27

THE EYE AS A CAMERA

Figure 261 illustrates the general organization of the eye, showing the optical system for focusing the image, and the *retina* on which the image is projected. The optical system is identical to that of a camera, and the retina corresponds to the photographic film. The retina translates the image into nerve impulses and transmits these into the brain through the optic nerve.

The outer envelope of the eye is a very strong bag composed mainly of a thick fibrous structure called the *sclera.* Anteriorly the sclera connects with the *cornea,* which is the clear part through which light enters the eye. The inside of the eye is filled with fluid, and an ovoid clear body called the *crystalline lens* is located approximately 2 mm. behind the cornea. The fluid anterior to the lens is almost pure extracellular fluid and is called *aqueous humor,* while that behind the lens contains a protein matrix that forms a gelatinous but clear fluid called the *vitreous humor.* Light passes

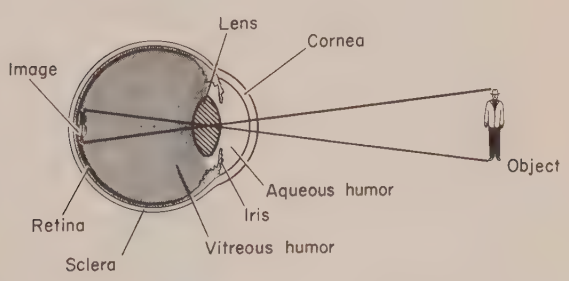


Figure 261. General structure of the eye, showing the function of the eye as a camera.

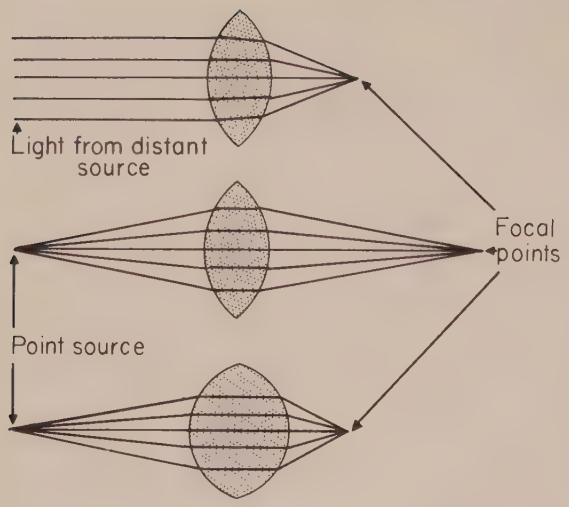


Figure 262. Focusing of parallel light rays and light rays from point sources by convex lenses.

first through the clear cornea, then through the aqueous humor, then the lens, and finally the vitreous humor before impinging on the retina.

The lens system of the eye. The lens system of the eye is composed of the cornea and the crystalline lens. Because the cornea is curved on its outside, light rays passing from the air into the cornea are *refracted* (bent) in the same way that any optical lens refracts rays. After the rays pass through the cornea and aqueous humor they strike the curved anterior surface of the crystalline lens, where still more bending occurs, and as they pass through the posterior surface of the lens the rays bend again. Thus, light rays are refracted, or bent, at three different interfaces in the eye. This is analogous to the refraction of light at the different surfaces in a compound lens system of a camera, for in the camera the light rays are also refracted at each interface between the lenses and the air.

Function of a convex lens in forming an image. Figure 262 shows the focusing of light rays from a distant source and also from two points near the lens. It will be noted that the light rays striking the outer edges of the convex lens are bent toward the center, while those that strike the lens exactly in the center pass on through without being bent. The reason for this is that the outer rays enter the substance of the lens at an angle, which causes refraction, while those striking the center enter the lens perpendicular to its surface, which does not cause refraction. The light rays from the edges of the lens angle inward to meet those passing through the center, and they all *focus* on a common point called the *focal point.*

Figure 263A shows the focusing of light rays from two different point sources. Note that the rays from the two lights are each focused to focal points on the opposite side of the lens directly in line with the lens center.

Figure 263B illustrates the focusing of light rays from different point sources on a human being's body. Those parts of the body that are very bright are point sources of light, whereas those parts that are dark represent the black spaces between the point sources of light. So far as the eyes are concerned, any object is a mosaic of point sources of light. Light from each source is focused by the lens, and the focal point of each is always directly in line with the center of the lens and the original source. Consequently, an inverted image is formed by the focal points, as shown in the figure.

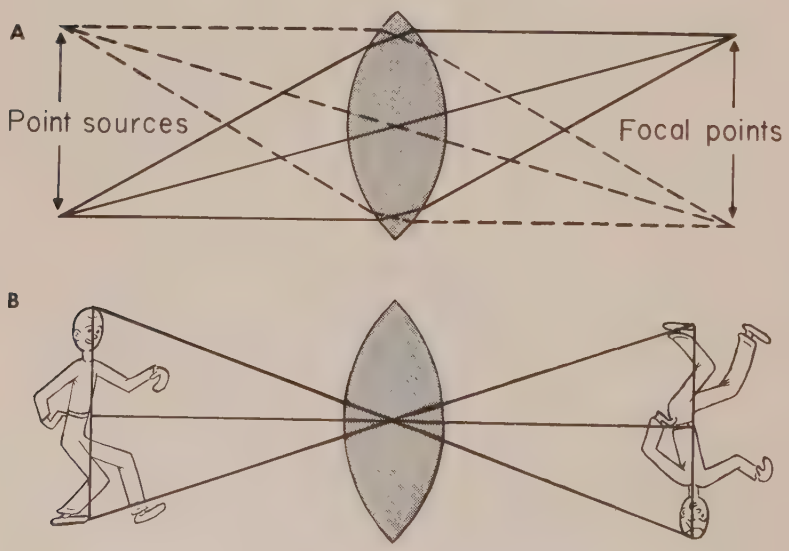


Figure 263. (A) Focusing of two points of light to two separate focal points. (B) Formation of an image by a convex lens focusing light rays from an object.

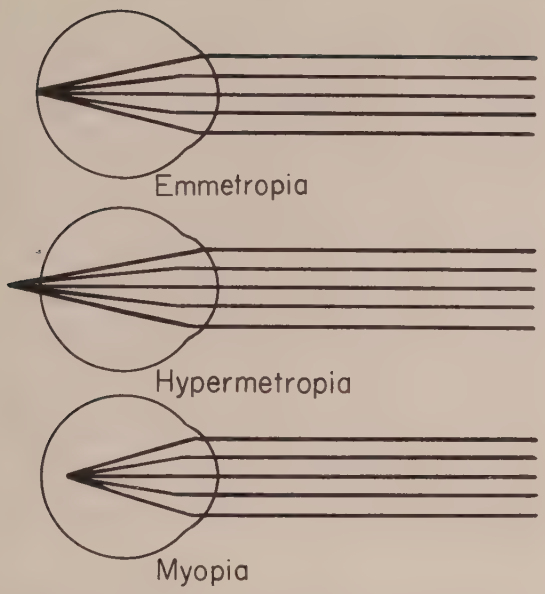


Figure 264. Focusing of light rays by the emmetropic eye, the hypermetropic eye, and the myopic eye.

Abnormalities of the Lens System

The normal eye focuses parallel light rays exactly on the retina. This normal focusing of the eye is called *emmetropia,* and it is illustrated at the top of Figure 264. However, three different abnormalities frequently occur to prevent focusing of light rays precisely on the retina. These are *hypermetropia, myopia,* and *astigmatism.*

Hypermetropia. Hypermetropia, or *far-sightedness,* is caused by failure of the lens to bend the light rays enough to bring them to a focal point on the retina, as shown in Figure 264. Instead, the light rays are still diffuse when they reach the retina, and thus vision is blurred. Hypermetropia is called far-sightedness because with this type of vision objects can be seen more clearly at a distance than near at hand.

Myopia. Myopia, which is also called *near-sightedness,* is caused by too strong a lens system for the distance of the retina behind the lens. That is, the light rays are focused before they reach the retina, and by the time they do reach the retina they have spread apart again, causing fuzziness of each point in the image.

Myopia is called near-sightedness because the myope can see objects near him with complete clarity, while not being able to focus any objects that are at a far distance. The reason for this is that the closer the object is to the eye, the more the light rays diverge outward so that the lens system cannot bend the rays inward as much as usual. If the object is close enough to the eye, the focal point will fall on the retina rather than in front of it, and the image will be completely clear.

Astigmatism. Astigmatism occurs when one of the components of the lens system becomes

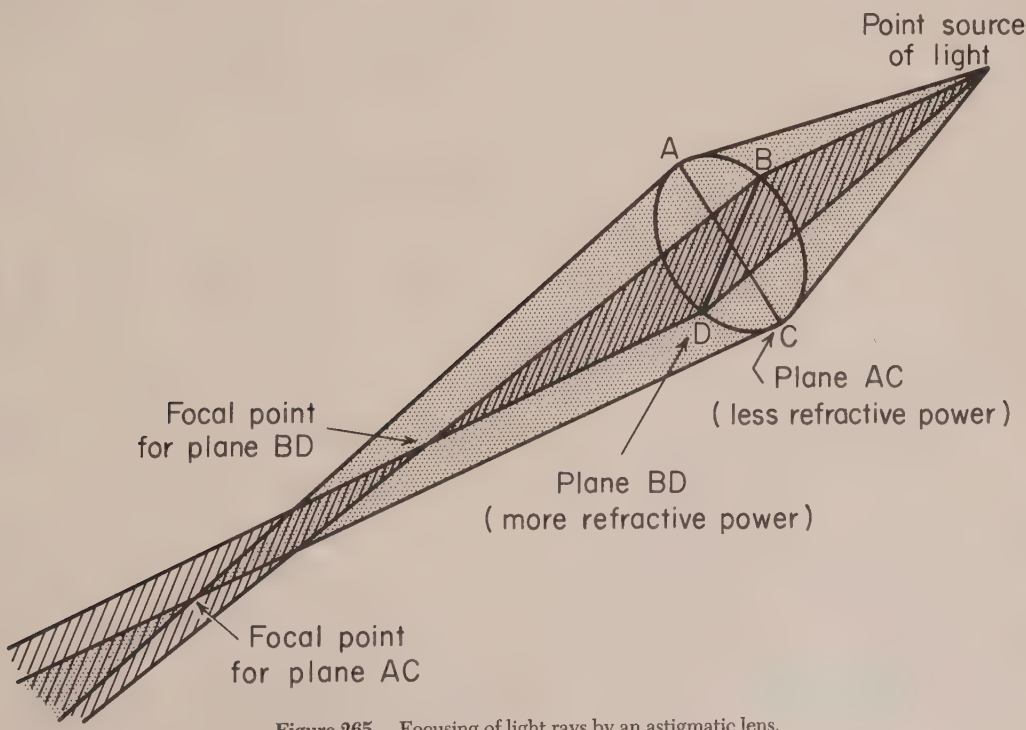


Figure 265. Focusing of light rays by an astigmatic lens.

egg-shaped rather than spherical. This effect is shown in Figure 265. Either the cornea or the crystalline lens becomes elongated in one direction in comparison with the other direction. Because the radius of curvature is greater in the elongated direction than in the short direction, the light rays entering the lens along this lengthened curvature are focused behind the retina, while those entering along the shortened curvature are focused in front of the retina. In other words, the eye is far-sighted for some of the light rays and near-sighted for the remainder. Therefore, the person with astigmatic eyes is unable to focus any object clearly, regardless of how far the object is away from the eyes, for when the near-sighted light rays are in focus the far-sighted ones are out of focus and vice versa.

Correction of the optical abnormalities of the eye. Glasses with properly prescribed lenses can be used to correct the abnormalities of the lens system of the eye. Glasses bend the light rays before they enter the eye in an appropriate manner to correct for the excess or deficient refractive power of the eye. Figure 266 shows the correction of myopia and hypermetropia. In the myopic individual the light rays normally focus in front of the retina. To prevent this, a concave lens is placed in front of the eye. This type of lens bends the light rays outward and, therefore, compensates for the excess inward bending of the myopic lens system. By prescribing the appropriate curvature to the concave lens, a myopic person's vision can be made completely normal.

In the hypermetropic eye the lens system normally fails to bend the light rays sufficiently. To correct this abnormality a convex lens is placed in front of the eye so that the light rays will be partially bent even before they reach the eye. With the aid of this preliminary convergence, the lens system of the eye can then bring the rays to a focal point on the retina.

The lens that must be used to correct astigmatism is somewhat more complicated than those used to correct either myopia or hypermetropia, for it must be fashioned with more curvature in one direction than the other. However, by prescribing a lens with precisely ground curvatures and by placing the curvatures in exactly the right direction in front of the eye, the abnormal refraction of light rays by each portion of the astigmatic eye can be corrected appropriately.

FUNCTION OF THE RETINA

Figure 267 illustrates the anatomy of the retina, showing that it is composed of many different layers. The light rays enter from the bottom of the figure and pass all the way through the retina to the top, finally striking the pigment layer. This pigment layer contains large quantities of a very black pigment called *melanin*, and the rays are completely stopped at this point. However, in their passage through the retina they impinge on the *rods* and *cones*, which are nerve receptors that are excited by light. These cells change the light energy into neuronal impulses that are transmitted into the brain.

The function of the pigment layer is to prevent

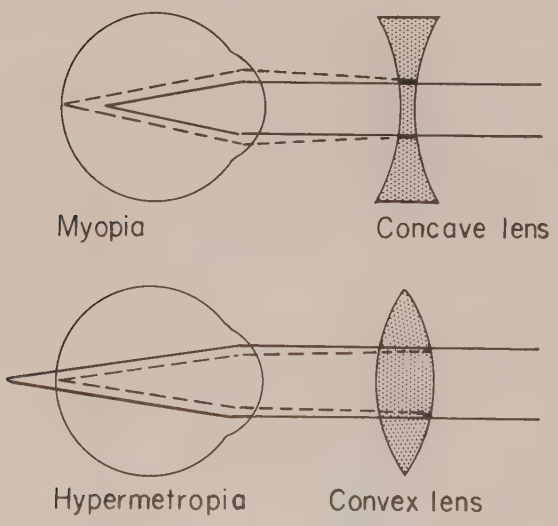


Figure 266. Correction of myopia by a concave lens and of hypermetropia by a convex lens.

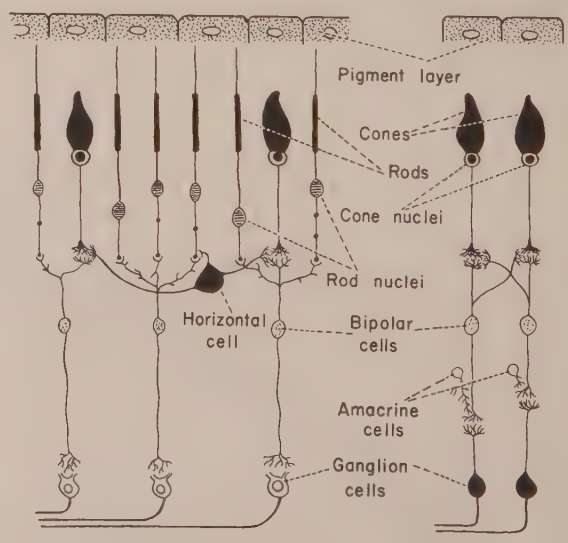


Figure 267. Functional anatomy of the retina.

light reflection throughout the eye. *Albino* persons, who are unable to manufacture melanin in any part of their bodies, have a complete lack of pigment in this layer of the retina. As a result, after passing through the retina the light rays are not absorbed, but instead are reflected so intensely in all directions through the eye that they cause all images to become bleached out with light. The albino's vision is usually about two to three times less acute than that of a normal person, and he is so blinded by bright sunlight that he must wear very dark glasses to see at all.

By far the greater number of light receptors in the retina are rods. Light of all colors stimulates the rods, while the cones are stimulated selectively by different colors. Therefore, the cones are responsible for color vision, in contradistinction to the rods, which provide only black and white vision.

The Chemistry of Vision by the Rods

Much more has been learned about the chemistry of vision in the rods than that in the cones. For this reason the method by which the rods change light energy into nerve impulses will be described. Figure 268 shows the chemical changes that occur in the rods. Vitamin A is the basic chemical utilized by both the rods and cones for synthesizing substances sensitive to light. On being absorbed into a rod, vitamin A is converted into a substance called *retinene.* This then combines with a protein in the rods called *scotopsin* to form a light-sensitive chemical, *rhodopsin.* If the eye is not being exposed to light energy, the concentration of rhodopsin builds up to an extremely high level.

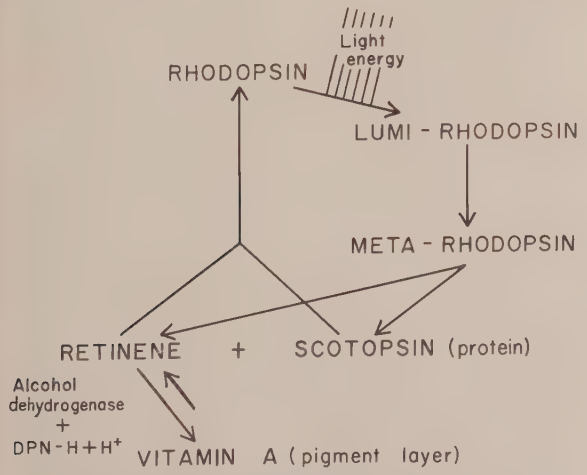


Figure 268. The retinene-rhodopsin chemical cycle responsible for light sensitivity of the rods.

When a rod is exposed to light energy some of the rhodopsin is changed immediately into *lumi-rhodopsin.* However, lumi-rhodopsin is a very unstable compound that can last in the retina only about a tenth of a second or more. It decays almost immediately into another substance called *meta-rhodopsin,* and this compound, which is also unstable, decays very rapidly into retinene and scotopsin. Thus, in effect, light energy breaks rhodopsin down into the substances from which the rhodopsin itself had been formed. In the process of splitting rhodopsin, the rods become excited, probably by ionic changes in the intracellular fluid when lumi-rhodopsin is formed. These ions last for only a split second. During this slight interval, nerve impulses are transmitted from the rod into the optic nerve and thence into the brain.

After rhodopsin has been decomposed by light energy, its decomposition products, retinene and scotopsin, are recombined slowly by the metabolic processes of the cell to form new rhodopsin. The new rhodopsin in turn can be utilized again to provide still more excitation of the rods. Thus, a continuous cycle occurs. Rhodopsin is being formed continually, and it is broken down by light energy to excite the rods.

Persistence of images and fusion of flickering images. Following a sudden flash of light that lasts only one millionth of a second, the eye sees an image of the light that lasts for approximately one-tenth second. The duration of the image is the length of time that the retina remains stimulated following the flash, and this presumably is as long as the lumi-rhodopsin remains in the rods.

The persistence of images in the retina allows flickering images to *fuse* when one views a moving picture or television screen. The moving picture flashes 16 to 30 pictures per second, and the television screen provides 60 pictures per second. The image on the retina persists from one picture to the next, which gives one the impression of seeing a continuous picture.

Light and dark adaptation. It is common experience to be almost totally blinded when first entering a very bright area from a darkened room, and also to be almost totally blinded when entering a darkened room from a brightly lighted area. The reason for this is that the sensitivity of the retina becomes temporarily out of adjustment with the intensity of the light. To discern the shape, the texture, and other qualities of an object, it is necessary to see both the bright and dark areas of the object at the same time. When the sensitivity of the retina is too great, all areas of the object will be bright enough to stimulate the rods a maximum amount, and, as a result, every-

thing—even dark objects—will seem to be completely white. This is the experience upon first coming into bright sunlight after seeing a movie. On the other hand, if the sensitivity of the retina is very slight when entering a darkened area, not even the bright portions of the image can stimulate the rods, and, as a result, no object can be seen. Fortunately, the retina automatically changes its sensitivity in proportion to the degree of light energy available. This phenomenon is called light and dark adaptation.

The mechanism of *light adaptation* can be explained by referring once again to Figure 268. When large quantities of light energy continually strike the rods, rhodopsin is broken into retinene and scotopsin almost as rapidly as it is formed, and, because rhodopsin formation is a relatively slow process, the total concentration of rhodopsin in the rods falls to a very low value as the person remains in the bright light. Therefore, the sensitivity of the retina soon becomes greatly depressed in bright light.

The mechanism of *dark adaptation* is essentially opposite to that of light adaptation. When the person enters a darkened room from a lighted area, the quantity of rhodopsin in his rods is at first very slight, and, as a result, he cannot see anything. Yet the amount of light energy in the darkened room is also very slight, which means that the rhodopsin being formed in the rods is utilized only very slowly. Therefore, the concentration of rhodopsin builds up during the ensuing minutes until it finally becomes high enough for even a very minute amount of light to stimulate the rods.

During dark adaptation, the sensitivity of the retina can increase as much as one thousand-fold in only a few minutes, and as much as one hundred thousand times in an hour or more. This effect is illustrated in Figure 269, which shows the retinal sensitivity increasing from an arbitrary value of 1 up to a value of 100,000 in one hour after the person has left a very bright area and moved into a completely darkened room. Then, on reentering the bright area, light adaptation occurs, and retinal sensitivity decreases from 100,000 back down to 1 in another 10 minutes, which is a more rapid process than dark adaptation.

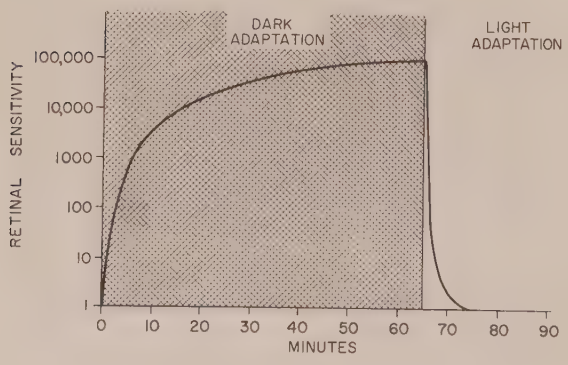


Figure 269. Dark and light adaptation.

Function of the Cones—Color Vision

The cones are different from the rods in several respects. First, they respond selectively to certain colors, some to one color and others to other colors. Second, cones are considerably less sensitive to light than are rods, for which reason they cannot provide vision in very dim light. Third, the cones are connected with the brain in such a way as to provide greater acuity of vision than the rods provide. That is, 10 to 100 rods usually connect with the same optic nerve fiber, which means that impulses transmitted by the rods to the brain do not necessarily originate from one very discrete point on the retina. On the other hand, usually only one and never more than a few cones are connected to the same optic nerve fiber.

Perception of different colors by the cones. The most important function of the cones is to detect colors, and this is accomplished by the ability of different cones to distinguish primary colors. In some lower animals as many as seven different types of cones, each sensitive to a different color, have been found. But it is believed, on the basis of psychological tests, that cones in the human eye are sensitive respectively to only three primary colors, blue, green, and red.

Figure 270 shows the so-called spectral sensitivities of the three different types of cones. To

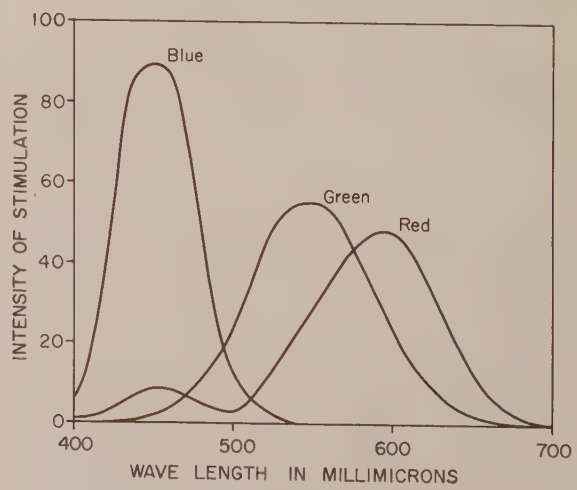


Figure 270. Spectral sensitivity curves for the blue, green, and red cones.

understand this chart, it is first necessary to know that light with a wave length of 400 millimicrons or less is violet in color; light at 450 millimicrons is blue; 500 millimicrons, green; 550 millimicrons, yellow; 580 millimicrons, orange; and above 600, red.

It is evident from Figure 270 that the blue cones are capable of responding to light of any wave length between 400 and 500 millimicrons. This means that the blue cones can receive violet, blue, and green light. However, they respond to the blue light with greater intensity than to any of the others.

The so-called green cones are capable of receiving light of wave lengths anywhere between 450 and 675 millimicrons. This means that the green cones can detect blue, green, yellow, orange, and red light. However, they respond to green light to a greater extent than to any of the others.

The red cones can detect light of all wave lengths, but they respond to a reddish-orange light more strongly than to any others.

Determination of the intermediate colors by the blue, green, and red cones. It is quite easy to understand how the blue cones determine that an object is blue, how the green cones determine that an object is green, and how the red cones determine that an object is red, but it is much more difficult to understand how these cones detect the intermediate colors between the three primary ones. This is accomplished by utilizing a combination of cones. For instance, yellow light stimulates the red and green cones approximately equally. When both of these types of cones are stimulated equally, the brain interprets the color as yellow. Also, when the red cones are stimulated about one and one-half times as strongly as the green cones, which occurs when light with a wave length of 580 millimicrons strikes the retina, the brain interprets the color as orange. If both the red and green cones are stimulated, but the green more than the red cones, the color is interpreted as a greenish-yellow. Likewise, when both the green and blue cones are stimulated, the color is interpreted as a bluish-green. If the intensity of stimulation of the green cones is greater than that of the blue cones, the color is interpreted as more green than blue, and when the blue cones are stimulated more than the green cones, the color is interpreted as more blue than green. Thus, by combining the degree of stimulation of the different cones, not only can the brain distinguish among the three primary colors but also among the colors having intermediate wave lengths.

The intensity of a color is determined by the number of nerve impulses transmitted to the brain by the cones. For instance, if a yellow color stimulates both the green and red cones and the number of impulses transmitted by each cone is only 10 per second, the intensity will be relatively weak, but if the number of impulses transmitted is 100 per second, the intensity will be strong. It is especially important to note that a change in intensity does not change the *ratio* between the degree of stimulation of the two types of cones. The brain interprets color on the basis of this ratio and not on the basis of the actual intensity of stimulation of each cone.

Color blindness. Color blindness can be understood very readily on the basis of Figure 270. Occasionally one of the three primary types of cones is lacking because of failure to inherit the appropriate gene for formation of the cone. The color genes are sex-linked and are found in the female sex chromosome. Since females have two of these chromosomes they almost never have a deficiency of a color gene, but because males have only one female chromosome, one or more of the color genes is absent in about 4 per cent of all males. For this reason, almost all color blind people are males.

If a person has complete lack of red cones, he is able to see green, yellow, orange, and red colors by use of his green cones. However, he is not able to distinguish satisfactorily between these particular colors because he has no red cones to contrast with the green ones. Likewise, if a person has a deficit of green cones, he is able to see all the colors, but he is not able to distinguish between green, yellow, orange, and red colors, because the green cones are not available to contrast with the red. Thus, loss of either the red or the green cones makes it difficult or impossible to distinguish between the colors of the longer wave lengths. This is called *red-green color blindness.* In very rare instances a person lacks blue cones, in which case he has difficulty distinguishing violet, blue, and green from each other. This type of color blindness is frequently called *blue weakness.*

NEURONAL CONNECTIONS OF THE RETINA WITH THE BRAIN

Figure 271 illustrates the connections of the retina with the brain, showing that the right halves of the two eyes are connected with the right visual cortex and the left halves with the left visual cortex. The *optic nerve* fibers from the nasal half of each retina cross in the *optic chiasm* to join the fibers from the temporal half of the opposite retina. Then the combined fibers pass backward through the *optic tract,* synapse in the *lateral geniculate body,* and finally spread through the *optic radiation* into the *visual cortex.*

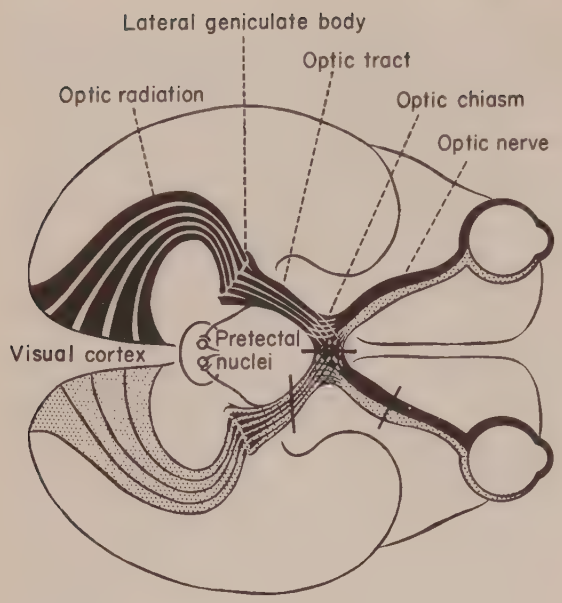


Figure 271. Optic pathways for transmission of visual signals from the retinae of the two eyes to the optic cortex. (From Polyak: *The Retina.* The University of Chicago Press.)

The fibers from each point in the retina pass to a particular point in the visual cortex, and for each different point in the retina a different point exists in the visual cortex. The image of an object on the retina is a mosaic of light and dark spots. These spots are transmitted in the form of nerve impulses to the visual cortex, where the image is reassembled in the form of stimulated and unstimulated neurons, the stimulated cells representing the light spots and the unstimulated cells the dark spots. In this way, the image on the retina is projected into the brain.

The portion of the visual cortex that receives the image from the retina is called the *primary visual cortex.* This area records light and darkness in the image, perhaps discriminates the different colors, and probably also distinguishes simple patterns of vision such as circles, squares, lines, and dots. Then the signals are transmitted into the *visual association areas,* where their meanings are interpreted. For instance, the picture of a word transmitted into the primary visual cortex would simply be a series of light and dark spots or at most simple letter patterns, but once this information is transmitted into the visual association areas the meaning of the word is determined. The function of the visual association areas as well as the other sensory association areas was discussed in detail in Chapter 23.

Fields of Vision

A means used to determine the extent of a person's normal vision and also to determine abnormalities of vision is to plot his *fields of vision.* The person is asked to close one eye and to look straight forward with the other eye. A small spot of light is moved above his central point of vision as far as he can see, then below as far as he can see, then to the right, to the left, and in all directions. In this way his ability to see in all areas away from his central-most point of vision is plotted.

Figure 272 shows the normal field of vision of the right eye. Far out to the right one can actually see objects at right angles (90 degrees) from the direction in which the eye is looking. To the left the nose is in the way, and the person can see objects only 50 degrees away from the central point of vision. Likewise, in the upward direction the orbital ridge is in the way, and in the downward direction the cheek bone is in the way. Were is not for these structures around the eye, the field of vision would be considerably greater.

In the field of vision is a *blind spot* caused by the *optic disk,* which is the point where the optic nerve enters the eyeball. At this point no rods or cones are present. The blind spot of each eye is located approximately 15 degrees to the temporal side of the central point of vision. However, the blind spots of the two eyes are on opposite sides in the respective fields of vision. The blind spots normally are not noticeable in one's vision.

Plotting the fields of vision provides a means for determining and locating damage in the retina or in the neuronal tracts from the retina to the

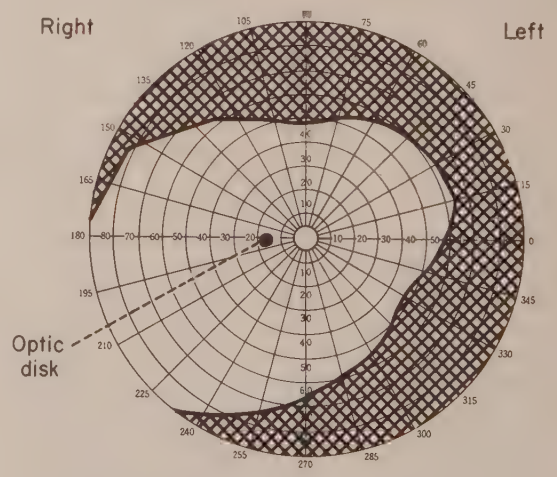


Figure 272. The visual field of the right eye.

brain. Referring back to Figure 271, it will be noted that three dark lines have been placed respectively across the right optic nerve, the optic chiasm, and the right optic tract. If the right optic nerve has been sectioned, the field of vision of the right eye will be zero, or, in other words, the eye will be completely blind. If the optic chiasm has been sectioned, the nasal half of each retina will be blind. This means that the *lateral* half of each visual field will be blind, because the lateral half of the field is registered by the nasal half of the retina. Finally, if the right optic tract is destroyed, the left halves of the visual fields of both eyes will be blind. Obviously, also, damage in any portion of the visual cortex or at any other point in the visual transmission system will cause loss of vision in respective areas in the fields of vision. The point in the eye or in the brain at which the damage has occurred can often be discerned from the pattern of visual loss.

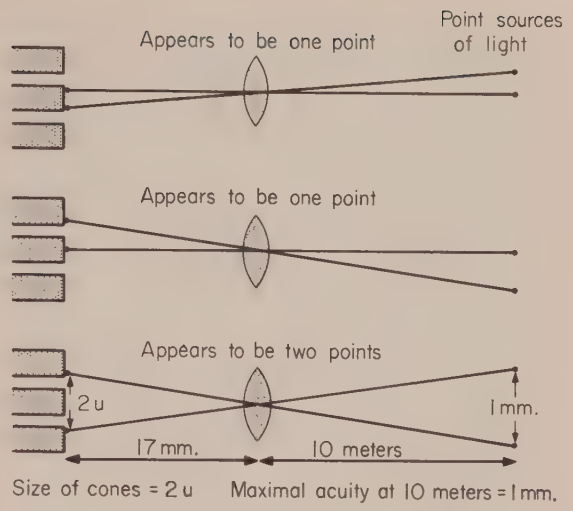


Figure 273. Focusing of points of light on cones in the fovea, illustrating the limits of maximal visual acuity.

Visual Acuity

Visual acuity means the degree of detail that the eye can discern in an image. The usual method for expressing visual acuity is to compare the person being tested with the normal person. If at a distance of 20 feet he can barely read letters of exactly the size that the normal person can barely read, he is said to have 20/20 vision, which is normal. If the letters must be as large as those that the normal person can read at a distance of 40 feet, his vision is said to be 20/40, or, in other words, his vision is one-half normal. If he can barely read letters at 20 feet distance that the normal person can read at 100 feet, his vision is 20/100, or, in other words, his vision is one-fifth normal. An occasional person has better than normal vision, so that he can read at 20 feet what the normal person can read only at 15 feet, in which case his vision is said to be 20/15.

Relationship of the cones to visual acuity. A person normally has very acute vision only in the central area of his visual field. The reason for this is that the central area of the retina, called the *fovea*, is especially adapted for acute vision. No rods are present in the fovea, and the cones there are considerably smaller in diameter than those in the peripheral portions of the retina. Also, the nerve fibers and blood vessels are all pulled to one side, so that light can pass with ease directly to the deep layers of the retina where the cones are located. Finally, and especially important, each of the cones of this region connects through an almost direct pathway with the brain so that the impulses from each cone do not become confused with impulses from other cones.

Effect of the optics of the eye on visual acuity. Figure 273 illustrates some of the optical factors that also influence the acuity of vision. It shows the images of point sources of light focused on the retina. If sources are so close together that both of the images focus on the same cone, then they will be interpreted as a single light rather than as two distinct lights. Likewise, if two lights are focused on two adjacent cones, they still will be interpreted as a single light, because a single light very frequently does stimulate two adjacent cones. Therefore, for the retina to interpret two points of light as separate lights, the images must impinge on cones separated by an unexcited cone between them. And because the diameter of each foveal cone is about 2 to 3 microns, the images of point sources of light must be about 2 to 3 microns apart before they can be seen as two separate points. Thus, in the lower part of Figure 273, it is shown that two points of light 10 meters away from the eye and 1 mm. apart can barely be discerned as two separate lights rather than as a single one by the normal eye. To express this another way, the maximal acuity of vision at 10 meters is 1 mm.

The normal lens system of the eye is sufficiently efficient so that visual acuity is determined mainly by the diameters of the cones in the fovea. When the optics of the eye are abnormal, visual acuity is determined instead by the ability of the lens system to focus the image on the retina. Consequently, a person with severe myopia, hypermetropia, or astigmatism may have a visual acuity

of 20/40, 20/100, 20/200, or so forth, depending on the degree of optical abnormality.

Depth Perception

The eyes determine the distance of an object by two principal means. These are, first, by the *size of the image* on the retina, and second, by the phenomenon called *parallax.* To determine the distance of an object from the eyes by the image size, the person must have had previous experience with the object and know its actual size. For example, our previous experience with other persons is that their average height is somewhere between 5 and 6 feet. Therefore, even when using one eye we can determine relatively accurately the distance of a person by the size of his image on the retina. If he is nearby, his image is very large, but, if he is far away, his image is very small.

The second means for determining the distance of an object, parallax, depends on slight differences between the shapes and positions of the images on the retinae of the two eyes. This effect, which is shown in Figure 274, occurs because the two eyes are set several inches apart. In this figure the two eyes are observing a ball near the eyes and a block at a distance. In the left eye the image of the block lies to the right of the image of the ball, but in the right eye, the image of the block lies to the left of the image of the ball. In other words, the images of these two objects are actually reversed on the retinae because of the angles from which the two eyes observe them, and the brain interprets their relative distances by the degree to which they are reversed. Obviously, this is an extreme example of parallax, but even when looking at a single object the images in the two eyes are slightly different. It must be noted particularly that this mechanism interprets only the *relative distances* of objects and not the actual distances. However, if the distance of one of the objects is known, that of the second object is also known. For practical considerations, a person always has his hands or other parts of his body to use as objects of known distance, which can be used as reference points for determining the distances of unknown objects.

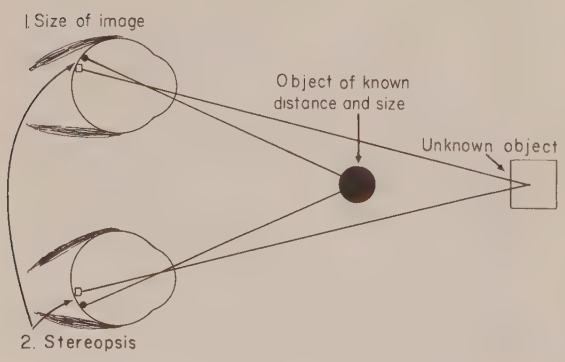


Figure 274. Mechanisms of depth perception.

NEUROGENIC CONTROL OF EYE MOVEMENTS

If the eyes are to function satisfactorily as a camera, their line of sight must be appropriately directed so that the portion of the image that is to be seen acutely will fall exactly on the fovea. In addition to this, the lens system of the eye must be focused for the distance of the object, and the pupil of the eye must be enlarged or contracted in proportion to the amount of light available, thus aiding in the light and dark adaptation of the eyes. It can be seen, then, that several different sets of eye muscles must be controlled very exactly to attain visual efficiency.

Positioning of the Eyes

Each eye is positioned by three different pairs of muscles, which are illustrated in Figure 275. One pair of these muscles is attached superiorly and inferiorly to the eyeball to move it up or down; another pair is attached horizontally to move the eye from side to side; and another pair is attached around the eyeball respectively on the bottom and top so that it can be rotated in either direction.

To control the eye movements, the visual association areas of the optic cortex must determine first whether or not the eyes are pointing toward the object; if not, impulses are transmitted through *oculomotor centers* in the brain stem to move the eyes in the appropriate direction. The visual association areas also determine whether or not the two eyes are receiving the same image on the corresponding portions of the two retinae —that is, whether or not the images are *fused* with each other. If not, one or both of the eyes are adjusted up or down, to one side or the other, or rotated so that the images will fuse. These minute adjustments of the eye positions are so important to vision that almost as much of the visual association areas in the brain is concerned with eye movements as with the interpretation of the meaning of visual signals.

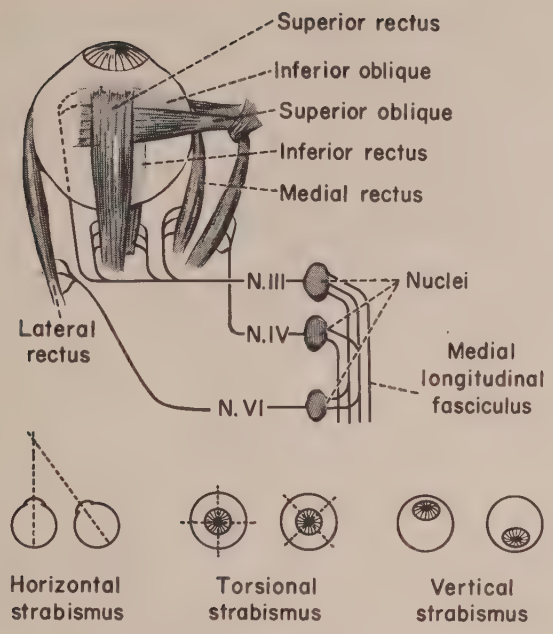


Figure 275. The extraocular muscles.

Movements of the eyes for following moving objects. Special mechanisms are also available in the visual association areas and cerebellum for allowing the eyes to follow moving objects. To achieve this, the eyes must move slowly in the same direction that the object is moving, neither falling behind nor getting ahead of the object. If the image does begin to fall behind or to move ahead, the visual association areas immediately send corrective signals to the eye muscles.

The cerebellum is involved in this mechanism because the object's course of movement must be predicted ahead of time; the predictive ability of the cerebellum allows the eyes to move along with the object rather than having to wait for the object to get to a new position before bringing the eyes along too.

Focusing of the Eyes

The eyes are kept focused on the visual field by changing the curvature of the crystalline lens as the field's distance changes. This is accomplished by the following neuromuscular mechanism:

Figure 276 shows the lens of the eye suspended from the sides of the eyeball by the suspensory ligaments, of which there are about seventy. The lens is actually an elastic ovoid structure having a clear envelope and clear viscous fluid in its center. When the ligaments are not pulling on the lens it assumes a round shape, but when the ligaments are tightened it flattens out. Normally, the ligaments are tight and the lens is flattened.

A smooth muscle called the *ciliary muscle* attaches to the suspensory ligaments of the lens where they connect with the eyeball. This muscle is composed of *meridional fibers* and *circular fibers*. The meridional fibers extend from the ends of the suspensory ligaments anteriorly to the sclerocorneal junction; when they contract they pull the ends of the suspensory ligaments forward and loosen them. This allows the lens to become thickened and the curvature of the lens to become greater, increasing the focusing power. The circular fibers extend all the way around the eye. When they contract they act as a sphincter, tightening progressively into a smaller and smaller circle, thus again loosening the suspensory ligaments so that the lens becomes still more convex and develops more focusing power.

Focusing, like positioning of the eyes, is accomplished by signals initiated in the visual association areas. When the image on the retina is out of focus, the fuzziness of the image initiates appropriate reactions to cause a change in the tension in the ciliary muscle. If the focus then becomes progressively worse, the visual association area immediately recognizes this and causes the tension in the ciliary muscle to begin changing in the other direction. As the focus becomes better and better it finally reaches a point at which the image has its greatest acuity. If the tension in the muscle keeps on changing, the eye goes out of focus in the opposite direction. Again the visual association area changes the direction of contraction of the ciliary muscle. Thus, the focusing mechanism keeps "searching" all the time for the image

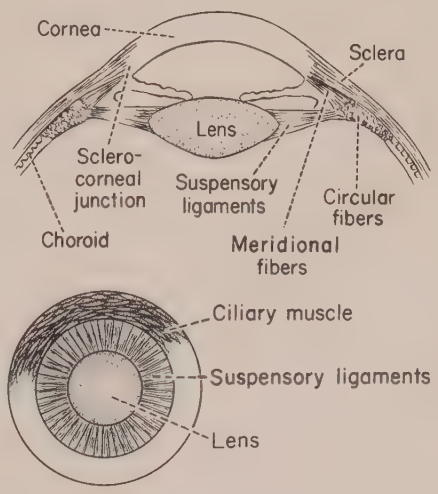


Figure 276. The focusing (accommodation) mechanism of the eye.

of maximal acuity, and even when the distance of the object from the eye changes tremendously, this searching mechanism can still find the appropriate focus within about one-fifth second.

Control of the Pupil

The pupil of the eye is the round opening of the iris through which light passes to the interior. The iris can constrict until the pupillary diameter is no greater than 1.5 mm., or it can relax until the diameter becomes 8 to 9 mm. Constriction of the pupil is caused by contraction of the *pupillary sphincter,* a circular muscle around the pupillary opening that is controlled by the parasympathetic nerves to the eye. Dilation of the pupil is caused by relaxation of the sphincter or by contraction of *radial muscle fibers* that extend from the edge of the pupil to the outer border of the iris; this muscle is controlled by the sympathetic nerves to the eye.

Because the amount of light that enters the eye is proportional to the area of the pupil and not to the diameter, the amount of light entering the eye can be varied approximately forty-fold. This provides a mechanism for light and dark adaptation in addition to the retinal mechanism for light and dark adaptation. Changes in pupillary size can occur in less than one second, in contrast with retinal adaptation which requires several minutes.

The size of the pupil is regulated by a *pupillary light reflex.* When the retina is stimulated by light, some of the resulting impulses, instead of passing all the way to the visual cortex, leave the optic tract as shown in Figure 271 and pass to the pretectal nuclei of the brain stem. From here signals are transmitted to the brain stem centers that control the muscles of the eye and finally back to the iris. When the light intensity in the eyes is great, the size of the pupillary opening diminishes. On the other hand, when the light intensity becomes slight, the opening of the pupil increases.

The pupillary light reflex is frequently absent in a person with syphilis of the central nervous system, for this disease has a special predilection for destroying the pretectal nuclei.

REFERENCES

Alder, F. H.: *Gifford's Textbook of Ophthalmology.* 7th Ed. Philadelphia, W. B. Saunders Company, 1962.

Alder, F. H.: *Physiology of the Eye: Clinical Application.* 3rd Ed. St. Louis, The C. V. Mosby Company, 1959.

Brindley, G. S.: *Physiology of the Retina and Visual Pathways.* Baltimore, The Williams & Wilkins Company, 1959.

Crescitelli, F.: Physiology of vision. *Ann. Rev. Physiol.* 22:525, 1960.

Dartnall, H. J. A., and Tansley, K.: Physiology of vision: retinal structure and visual pigments. *Ann. Rev. Physiol.* 25:433, 1963.

Duke-Elder, S.: *Textbook of Ophthalmology.* Vols. I-VI. St. Louis, The C. V. Mosby Company, 1932, 1938, 1941, 1949, 1952, 1954.

Granit, R.: *Receptors and Sensory Perception.* New Haven, Yale University Press, 1955.

Linksz, A.: *Physiology of the Eye.* Vol. II. Vision. New York, Grune & Stratton, Inc., 1952.

Lyle, D. J.: *Neuro-ophthalmology.* 2nd Ed. Springfield, Ill., Charles C Thomas, 1954.

Polyak, S. L.: *The Retina.* Chicago, The University of Chicago Press, 1941.

Wald, G.: The photoreceptor process in vision. *Handbook of Physiology* 1:671, 1959.

Willmer, E. N.: The physiology of vision. *Ann. Rev. Physiol.* 17:339, 1955.

HEARING, TASTE, AND SMELL

28

The function of the ear is to convert sound energy into nerve impulses. Figure 277 illustrates the general organization of the ear, showing the *external ear,* the *auditory canal,* the *tympanic membrane,* and the *ossicular system,* composed of the *malleus, incus,* and *stapes,* that transmits sound energy into the *cochlea.* The cochlea is the portion of the inner ear that actually converts sound into nerve impulses, while the remainder of the anatomic structures simply transmits sound waves from the air to the cochlea.

TRANSMISSION OF SOUND TO THE INNER EAR

Characteristics of sound and sound waves. Sound is caused by compression waves traveling through the air. The transmitter of the sound, whether it be another person's voice, a radio speaker, or some noise-making device, creates the sound by alternately compressing the air and then relaxing the compression. For instance, a vibrating violin string creates sound by moving back and forth. When the string moves forward it compresses the air, and when it moves backward it decreases the amount of compression even below normal. This alternate compression and evacuation of the air produces sound.

Sound waves travel through the air in very much the same way that waves travel over the surface of water. Compression of the air adjacent

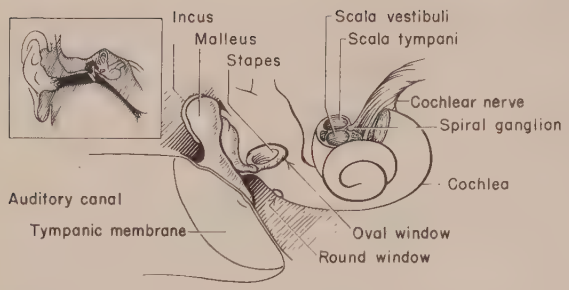


Figure 277. General organization of the ear, showing the external ear, the ossicular system, and the cochlea.

to a violin string builds up extra pressure in this region, and this in turn causes the air a little further away to become compressed. The pressure in this second region then compresses the air still further away, and this process continues on and on until the wave finally reaches the ear.

Function of the Tympanic Membrane (Tympanum) and Ossicles

When sound waves strike the tympanic membrane, the alternate compression and evacuation of the air adjacent to the membrane cause it to move backward and forward. The center of this membrane is connected to the handle of the *malleus;* this in turn is connected to the *incus,* and the incus to the *stapes.* These bones are called the *ossicles.* They are suspended in the *middle ear* by ligaments, so that they can rock back and forth. Movement of the handle of the malleus, therefore, causes the stapes also to move back and forth against the *oval window* of the cochlea, thus transmitting the sound into the cochlear fluid.

Transformation of the pressure of the sound waves by the ossicular system. If sound waves were applied directly to the oval window, they would not have enough pressure to move the fluid in the cochlea backward and forward to produce adequate hearing, because fluid has many times more inertia than air, and a correspondingly greater amount of pressure is required to cause movement in fluid. The tympanic membrane and the ossicular system transform the pressure of the sound waves into a usable form by the following means:

The sound waves are collected by a very large membrane, the tympanic membrane, the area of which is approximately 70 square millimeters, or 22 times that of the oval window which has an area of only 3.2 square millimeters. Therefore, twenty-two times as much sound energy is collected as could be collected by the oval window alone, and all of this is transmitted through the ossicles to the oval window. Thus, the pressure of movement of the foot of the stapes is increased to about 22 times that which could be effected by applying sound waves directly to the oval window. This pressure is now sufficient to move the fluid of the cochlea backward and forward.

Transmission of Sound into the Cochlea by the Bones of the Head

The cochlea lies in a bony chamber inside the temporal bone of the skull. Any vibration of the skull also causes vibration of the fluid in the cochlea. Ordinarily, sound waves in the air cause almost no vibration in the skull bones, but clicking of the teeth or holding vibrating devices such as a tuning fork or special sound vibrators against the skull can cause bone vibrations. In this way, instead of the fluid vibrating inside the cochlea, the cochlea vibrates around the fluid, allowing the cochlea to interpret the sound as well as if it had entered via the tympanic membrane and ossicles.

FUNCTION OF THE COCHLEA

Physiologic anatomy of the cochlea. The cochlea is a membranous device formed of coiled tubes. The outside appearance of the cochlea is illustrated in Figure 277, and a cross-section of its structure is shown in Figure 278. If the coiled

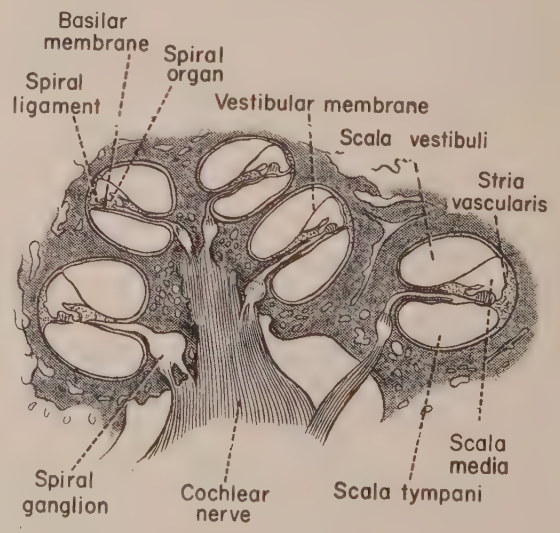


Figure 278. Cross-section of the cochlea, showing its coiled tubes composed of the scala vestibuli, the scala tympani, and the scala media. The cochlear nerve leading from the basilar membrane is also illustrated. (From Goss: *Gray's Anatomy of the Human Body,* Lea & Febiger.)

tubes were stretched out, their total length would be about 3.5 cm.

The cross-sectional diagram of Figure 278 shows that the cochlea is actually composed of three separate tubes, lying side by side, called the *scala vestibuli,* the *scala tympani,* and the *scala media.* All of these tubes are filled with fluid, and they are separated from each other by membranes. The membrane between the scala vestibuli and scala media is so thin that it never obstructs the passage of sound waves. Its function is simply to separate the fluid of the scala media from that of the scala vestibuli. The two fluids have separate origins and the difference between them is believed to be important for proper operation of the sound receptor cells. On the other hand, the membrane separating the scala media from the scala tympani, called the *basilar membrane,* is a very strong structure that does impede the sound waves. It is supported by about 25,000 ligamentous fibers, called *basilar fibers,* that stretch across the membrane. Located on the surface of the basilar membrane are the sound receptor cells called the *hair cells.*

Figure 279 depicts the cochlea diagrammatically after it has been uncoiled. In this figure the membrane between the scala vestibuli and the scala media has been eliminated because it does not affect sound transmission in the cochlea in any way. So far as the transmission of sound is concerned, the cochlea is only two separate tubes rather than three, and the two tubes are separated by the basilar membrane. The basilar fibers near the oval and round windows are short, but they become progressively longer, like the strings of a harp, farther and farther up the cochlea, until at the tip of the cochlea the fibers are about two and one-half times as long as those near the base.

Resonance of Sound in the Cochlea

Conduction of sound in the cochlea. As each sound vibration enters the cochlea, the oval window at first moves inward and pushes the fluid of the scala vestibuli up into the cochlea, as indicated by the arrows in Figure 279. The sudden pressure in the scala vestibuli bulges the basilar membrane into the scala tympani; this pushes fluid in this chamber toward the round window and causes it to bulge outward. Then, as the sound vibration causes the stapes to move backward, the procedure is reversed, with the fluid moving in the opposite direction along the same pathway and the basilar membrane bulging now into the scala vestibuli.

Resonance in the cochlea. A phenomenon

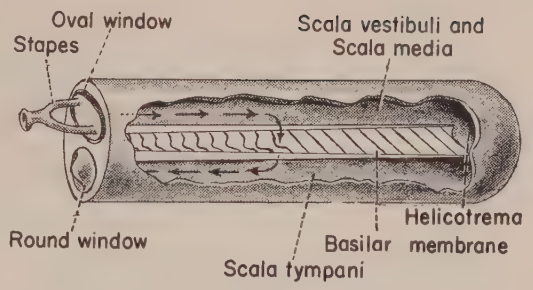


Figure 279. Movement of fluid in the cochlea when the stapes is thrust forward.

called *resonance* occurs in the cochlea to allow each sound frequency to vibrate a different section of the basilar membrane. These vibrations are similar to those that occur in many musical instruments and can be explained as follows: When the string of a violin is pulled to one side is becomes stretched a little more than usual, and this stretch makes the string then move back in the other direction. However, as it moves it builds up momentum and, therefore, does not stop moving when it reaches its normal straight position. The momentum causes the string to become stretched once again but in the opposite direction this time, and the string moves back in the first direction. The cycle continues over and over again so that once the string starts vibrating it continues.

Two factors determine the frequency at which a string will vibrate. First, the greater the *tension* developed by the string, the more rapidly it turns around and moves in the opposite direction, and the higher is the frequency of vibration. The second factor is the *mass* of the string. The greater its mass, the greater is the momentum developed during vibration, and the longer it will take for the string to change direction of movement; therefore, the lower will be the sound frequency. These same two factors, mass and elastic tension, apply to resonance in the cochlea. The vibrating mass in the cochlea is the fluid that vibrates back and forth between the oval and round windows, while the elastic tension is mainly the tension developed by the basilar fibers.

When *high frequency* sound enters the oval window, the sound wave travels along the basilar membrane only a short distance, as shown in Figure 280A, before the point of resonance is reached. The basilar membrane at this point vibrates "in tune" with the frequency of the sound. When a *medium frequency* sound enters the oval window, the wave travels much further along the basilar membrane before the area of resonance is reached, as illustrated in Figure 280B. And, finally, a *low-frequency* sound wave travels almost all the way to the end of the membrane before

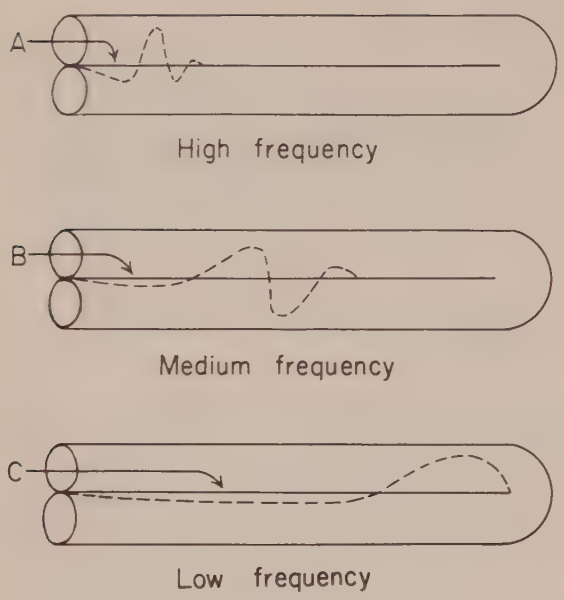


Figure 280. Diagrammatic representation of sound waves traveling along the basilar membrane and resonating at different points for high, medium, and low frequency sounds.

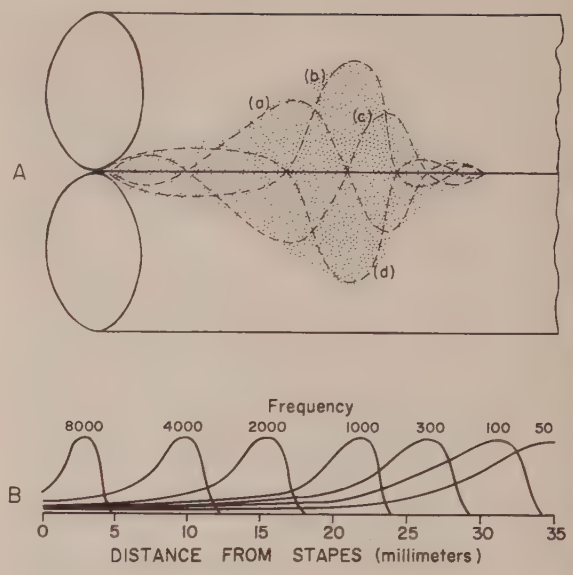


Figure 281. (A) Amplitude pattern of vibration of the basilar membrane for a medium frequency sound. (B) Amplitude patterns for sounds of frequencies between 50 and 8000 per second, showing resonance at different points for the different frequencies.

it reaches its resonance point, as illustrated in Figure 280C.

High frequency sound waves resonate near the base of the cochlea for the following reasons: As the stapes moves forward at the oval window, the fluid in the scala vestibuli also moves forward; this causes the basilar membrane to bulge toward the scala tympani and the fluid in this chamber to move backward toward the round window. Then as the stapes moves backward, exactly the opposite effects take place. Thus, the mass of fluid between the oval and round windows and the point of vibration of the basilar membrane is the mass of the vibrating system.

The mass of fluid between the windows and the basal regions of the basilar membrane is very slight. However, the basilar fibers, which provide the elastic tension for the vibrating system, have far greater rigidity at the basal end of the membrane than toward its tip. The combined effect of these two factors, the low mass and the greater rigidity, causes the basilar membrane to resonate at very high frequencies near its base. On the other hand, when a sound wave travels all the way to the tip of the basilar membrane, the mass of fluid that vibrates is very great. Also, the basilar fibers are longer and less rigid near the tip of the cochlea. The combined effect of these two factors, the great mass and the low rigidity, makes the basilar membrane near this end resonate at a very low frequency, and intermediate frequencies resonate at intermediate points along the membrane. Figure 281A illustrates the amplitude of vibration of different parts of the basilar membrane for a medium frequency sound wave, and Figure 281B represents the degree of vibration of the membrane for several different sound frequencies from very low to very high.

Determination of pitch by the cochlea. Located on the surface of all the basilar fibers are special nerve receptors, called *hair cells,* that are stimulated when the basilar fibers vibrate back and forth. These are shown in Figure 282. When the hair cells near the base of the cochlea are stimulated, the brain interprets the pitch of the sound as that of a very high frequency. When the hair cells in mid-cochlea are stimulated, the brain interprets the sound as one of intermediate pitch, and stimulation of those at the tip of the cochlea is interpreted as low pitch. It can be seen, therefore, that the pitch of a sound is determined by the point in the cochlea at which the basilar membrane vibrates.

Determination of loudness of a sound. The loudness of a sound is determined by the intensity of movement of the basilar fibers. The greater the displacement back and forth, the more intensely are the hair cells stimulated, and the greater are the number of impulses transmitted into the brain to indicate the degree of loudness. For instance, if a hair cell near the base of the cochlea transmits only one impulse per second, the frequency of the sound will still be interpreted as one of very high pitch, but the loudness of the

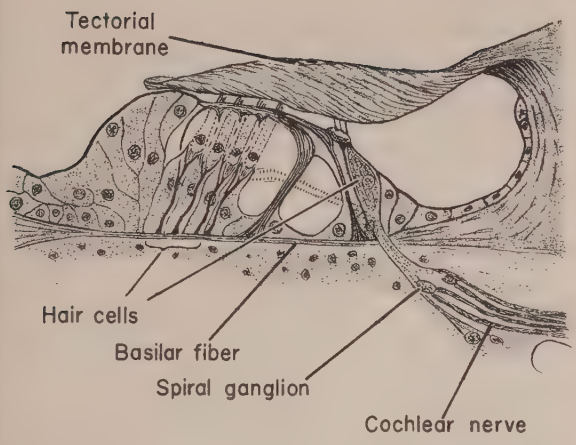


Figure 282. The organ of Corti, showing the hair cells and the tectorial membrane pressing against the projecting hairs. (Modified from Bloom and Fawcett: *A Textbook of Histology.* 8th Ed.)

sound will be almost zero. If the same hair cell is stimulated 1000 times per second, the pitch of the sound will remain the same, but the loudness will be extreme.

Conversion of Sound Energy into Nerve Impulses

The method by which vibration of the basilar fibers converts sound energy into nerve impulses is not completely understood, but experiments have shown that vibration of the basilar fibers causes the hair cells to generate electrical potentials. Movement of the basilar membrane in one direction causes the potential above the membrane to become positive and below negative. Then, when the basilar fibers vibrate in the opposite direction, the potential shifts polarity. Therefore, the method by which the hair cells excite the cochlear nerve endings is presumably by some electrical transfer from the hair cells to the filamentous endings of the cochlear nerve, which is illustrated in Figures 278 and 282.

The hair cells of the organ of Corti are almost identical with the hair cells of the vestibular apparatus for maintaining equilibrium, which were described in Chapter 25. It is known that the hair cells in the latter apparatus are stimulated by bending the hairy projections that extend from the ends of the cells. Therefore, it can be presumed that the hair cells of the cochlea are also stimulated by bending their hairs. A suggested mechanism by which this could occur involves the tectorial membrane, which is shown in Figure 282. Vibration of the basilar membrane causes the hair cells lying on its upper surface to rock back and forth; this is believed to brush the projecting hairs against the tectorial membrane. Movement of the hairs excites the cells and generates impulses in the small filamentous cochlear nerve endings that entwine the hair cells.

TRANSMISSION OF SOUND INTO THE CENTRAL NERVOUS SYSTEM

Figure 283 shows the transmission of sound impulses from the cochlea into the central nervous system. After passing through the cochlear nerve, the impulses are transmitted through five different structures, which may be listed as follows: (1) the *dorsal* and *ventral cochlear nuclei,* (2) the *trapezoid body* and *superior olivary nucleus,* (3) the *inferior colliculus,* (4) the *medial geniculate body,* and (5) the *auditory cortex.*

Function of lower auditory centers. The function of the lower auditory centers is not well understood, but some of them are probably responsible for localization of the direction from which sound is coming. For instance, if sound comes from the left, it reaches the left ear before it reaches the right. If the sound is coming from directly in front, it reaches both ears simultaneously, but as its direction changes from directly ahead to one side, the time lag becomes progressively greater. The time lag is presumably interpreted in some of the lower auditory centers, to determine the direction from which the sound is coming.

Another function of the auditory centers in the brain stem is the production of rapid motions of the head, of the eyes, or even of the entire body in response to auditory signals. Some of these signals pass into the cerebellum and bulboreticular formation to alter a person's equilibrium, to make his head jerk to one side, or to make his eyes roll in one direction. Some of these rapid responses can occur even though the cerebral portions of the auditory system have been destroyed.

Figure 283 shows that the auditory signals from each ear are transmitted approximately equally into the auditory pathways on both sides of the brain stem and cerebral cortex. Therefore, destruction of one of the pathways will not greatly affect one's hearing ability in either ear.

Function of the auditory cortex. The function of the auditory cortex has already been discussed in Chapter 23. In brief, the *primary auditory cortex* is located in the middle of the superior gyrus of the temporal lobe. This area receives the sound signals and interprets them

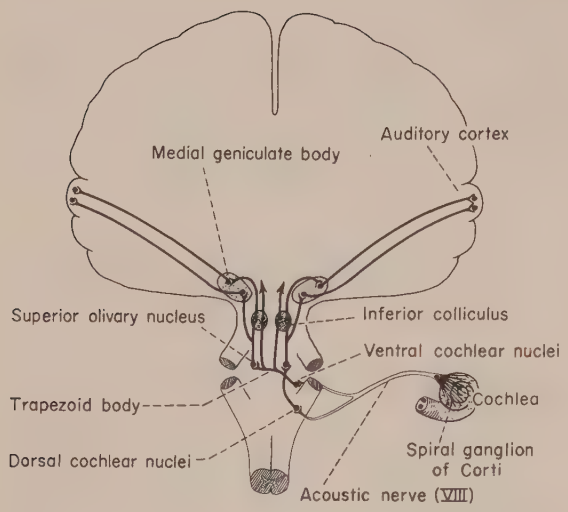


Figure 283. Pathways for transmission of sound impulses from the cochlea into the central nervous system.

as different sounds. However, the signals must then be transmitted into surrounding *auditory association areas* before their meanings become clear. Finally, the signals are transmitted into the *common integrative center* of the cortex, where the overall meaning of all combined auditory, visual, and other types of sensation is determined.

Destruction of one auditory cortex reduces hardly at all one's ability to hear, but destruction of both cortices greatly depresses the hearing. It is especially interesting, though, that even with both auditory cortices destroyed, an animal is still capable of hearing sounds of extreme intensity. It seems that the medial geniculate bodies of the thalamus are the regions that receive these sounds, for destruction of these areas causes total deafness. This illustrates again that many sensory functions of the brain can be performed by the thalamus independently of the cerebral cortex.

DEAFNESS AND HEARING TESTS

Any damage to the ossicular system, the cochlea, the cochlear nerve, or to the pathways for transmission of sound to the auditory cortex can cause partial or total deafness. The types of deafness are generally separated into two categories, *conduction deafness* and *nerve deafness*. The term conduction deafness means deafness caused by failure of sound waves to be conducted from the tympanic membrane through the ossicular system into the cochlea. Nerve deafness, on the other hand, means failure of the sound impulses to reach the auditory cortex because of damage to the cochlea itself or to the neurogenic transmission system for sound.

Probably the most common cause of *conduction deafness* is repeated blockage of the *auditory tube*. This tube connects the middle ear with the nose. Its function is to keep the pressure inside the middle ear equal to the pressure of the air, so that no pressure difference will exist between the two sides of the tympanic membrane. If the auditory tube becomes plugged because of a cold, allergic swelling of the nasal membranes, or some other cause, then the air in the middle ear becomes absorbed, and in its place a serous fluid collects. Also, the tympanic membrane is pulled inward because of lowered pressure in the middle ear. Fibroblasts tend to grow into the serous fluid and cause fibrous tissue between the ossicles and the walls of the middle ear. If this process continues long enough, the ossicles finally become so firmly bound to the walls of the middle ear that sound conduction into the cochlea becomes almost nil.

Nerve deafness is characteristic of old age; almost all older people normally develop at least some degree of this type of deafness, especially for very high frequency sounds. This is probably caused by the aging process in the cochlea itself, though the auditory pathways in the central nervous system possibly degenerate to some extent also.

Hearing Tests

Almost any type of sound instrument can be used to check a person's hearing; the most common of these for many years has been the *tuning fork*. After striking a tuning fork the sound can usually be heard for about 30 seconds if the fork is held near the normal ear. If the person has conduction deafness, the ear is unable to hear the sound, but placing the butt of the fork against the skull will allow transmission of vibrations to the cochlea via the bones of the skull. If the cochlea is still functioning properly, the sound is then heard even though it could not be heard by air conduction. On the other hand, if the person has nerve deafness, he is unable to hear the tuning fork either by air or bone conduction.

The audiometer. In recent years a special sound emitter, called an audiometer, has been used to measure the degree of deafness. This is an electronic apparatus capable of generating sound of all frequencies in an earphone or in a vibrator placed against the bone of the skull. The apparatus is calibrated so that the zero mark corresponds to the intensity of sound required for

a normal person barely to hear it. If the person is deaf for sound of a particular frequency, the sound intensity must be increased far above the zero level before he can hear it, and the amount of extra sound energy that must be added is said to be the *hearing loss* for that particular frequency.

Figure 284 shows an *audiogram* from a person who has conduction deafness. Approximately 40 to 60 decibels of extra sound energy had to be transmitted into the earphone at each frequency for the person to hear the sound. However, when the skull vibrator was used instead of the earphone, no extra energy was required; indeed, at the high frequencies the hearing was even better than normal. This illustrates that bone conduction was normal, which means that the cochlea and auditory pathways were also normal. However, conduction of sound through the ossicular system was greatly impaired.

The decibel system for expressing sound energy is a logarithmic scale rather than a linear one. As a result of this, a change of 10 decibels is actually a 10-fold change in energy, 20 decibels is a 100-fold change, 30 decibels is 1000-fold, and 60 decibels is 1,000,000-fold. For example, in the audiogram of Figure 284, the hearing loss at most frequencies is about 60 decibels, which means that this deaf person actually had a hearing loss of about 1,000,000-fold.

Sounds that the normal ear hears frequently vary in intensity a hundred million-fold or more. For instance, the intensity of sound in a very noisy factory is about one million times as great as that of a quiet whisper. Therefore, a person with 60 decibels hearing loss—a one million-fold loss—can still hear sounds of very strong intensity.

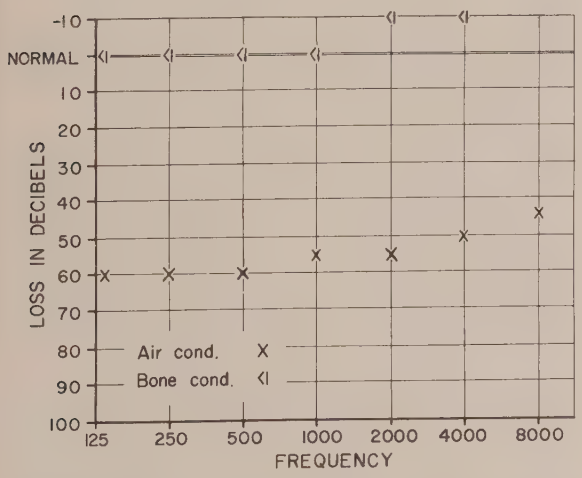


Figure 284. An audiogram from the ear of a person with conduction deafness.

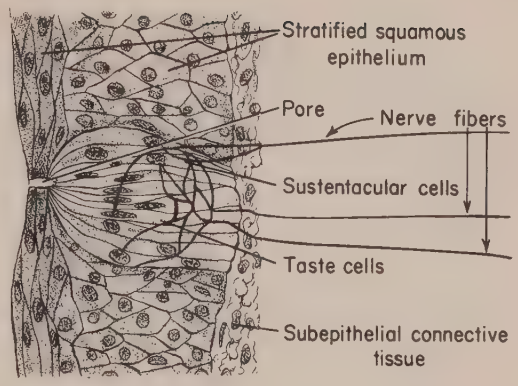


Figure 285. A taste bud. (Modified from Bloom and Fawcett: *A Textbook of Histology.* 8th Ed.)

TASTE

The senses of taste and smell are called the *chemical senses* because their receptors are excited by chemical stimuli. The taste receptors are excited by chemical substances in the food that we eat, while the smell receptors are excited by chemical substances in the air.

The Taste Bud

The sensory receptor for taste is the *taste bud,* shown in Figure 285. This is composed of epithelioid type cells arranged to form a small *pore* in the mucous membrane of the mouth. A single nerve fiber passes from each taste bud into the taste nerves and then to the central nervous system.

Before a substance can be tasted it must be dissolved in the fluid of the mouth, and it must diffuse through the taste pore to the tips of the taste cells. Therefore, highly diffusible substances, such as salts or other small molecular compounds, generally cause greater degrees of taste than do less diffusible substances, such as proteins or other substances of very large molecular size.

The four primary sensations of taste. There are four different types of taste buds, each of which detects a different type of taste sensation. These four sensations are called the *primary* taste sensations, and they are (1) salty, (2) sweet, (3) bitter, and (4) sour.

Almost any electrolyte will stimulate the salty buds. These buds determine the proportion of salts and other electrolytic substances present in food. The most familiar salt that stimulates these buds is sodium chloride, common table salt. Therefore, we normally associate the function of

these buds almost entirely with the content of table salt in food.

The sweet taste buds detect the amount of sugars in food. This is one of the means by which animals determine whether or not fruits are ripe and whether or not unknown foods are nutritious. Essentially all wild foods that are sweet are safe to eat and contain a considerable amount of nutrition.

The bitter taste buds provide a protective function, for they detect principally the poisons in wild plants. The alkaloidal compounds of wild herbs, for instance, are among the poisons detected by the bitter taste buds. These same alkaloids, when given in appropriate quantities, are frequently very valuable drugs. Many drugs such as quinine, though of very bitter taste, have a beneficial effect when used properly, and yet can be lethal when used improperly.

The sour taste buds detect the degree of acidity of food—that is, they detect the concentration of hydrogen ion in the mouth. If the degree of acidity is slight, such as that of dilute vinegar, the food is usually very palatable; if the acidity is extremely strong, the taste is very unpleasant, and the food is rejected.

Regulation of the Diet by the Taste Sensations

The taste sensations obviously help to regulate the diet. For example, the sweet taste is usually pleasant, causing an animal to choose foods that are sweet. On the other hand, the bitter sensation is always unpleasant, causing bitter foods that are often poisonous to be rejected. The sour taste is sometimes pleasant and sometimes unpleasant, and, likewise, the salty taste is sometimes pleasant and sometimes unpleasant. The pleasantness of these types of taste is often determined by the momentary state of nutrition of the body. If a person has been long without salt, then, for reasons which are not yet understood, the salty sensation becomes extremely pleasant. If a person has been eating an excess of salt, the salty taste is very unpleasant. The same is true for the sour taste and to a less extent for the sweet taste. In this way the quality of the diet is automatically varied in accord with the needs of the body. That is, lack of a particular type of nutrient often intensifies one or more of the taste sensations and causes the person to choose foods having a taste characteristic of the deficient nutrient.

Importance of smell to the sensation of taste. Much of what we call taste is actually smell, because foods entering the mouth give off odors that spread into the nose. Often, a person who has a cold states that he has lost his sense of taste, but on testing for the four primary sensations of taste they are all still present.

The smell sensations, which are discussed below, function along with taste sensations to help control the appetite and the intake of food.

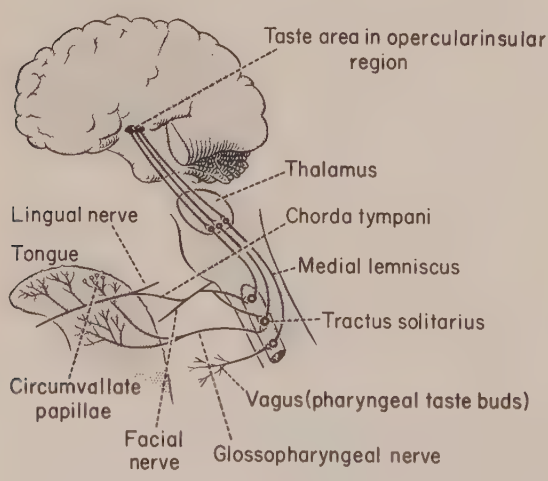


Figure 286. Transmission of taste impulses into the central nervous system.

Transmission of Taste Sensations to the Central Nervous System

Figure 286 shows the transmission of taste sensations into the brain stem and then into the cerebral cortex. Taste sensations pass from the mouth to the *tractus solitarius* located in the medulla. Then signals are transmitted to the *thalamus,* and from there to the *primary taste cortex* of the opercularinsular region. From here the taste sensations are transmitted into surrounding taste association areas, and finally into the common integrative region, which integrates all sensations.

Taste reflexes. One of the functions of the taste apparatus is to provide reflexes to the salivary glands of the mouth. To do this, impulses are transmitted from the tractus solitarius into nearby nuclei that control secretion by the parotid, submaxillary, and other salivary glands. When food is eaten, the quality of the taste sensations, operating through these reflexes, determines whether the output of saliva will be great or slight.

SMELL

The sense of smell is vested in two small areas, called the *olfactory epithelium,* that lie in the

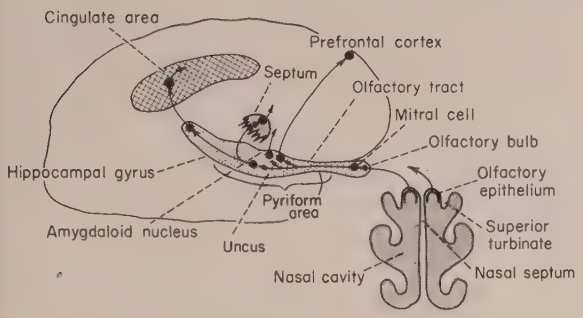


Figure 287. Organization of the olfactory system, and pathways for transmission of olfactory impulses into the central nervous system.

upper reaches of the nasal cavity. The location of these areas is illustrated in Figure 287, which shows to the right a cross-section of the air passages of the nose, and also shows the olfactory epithelium connected with the nervous system.

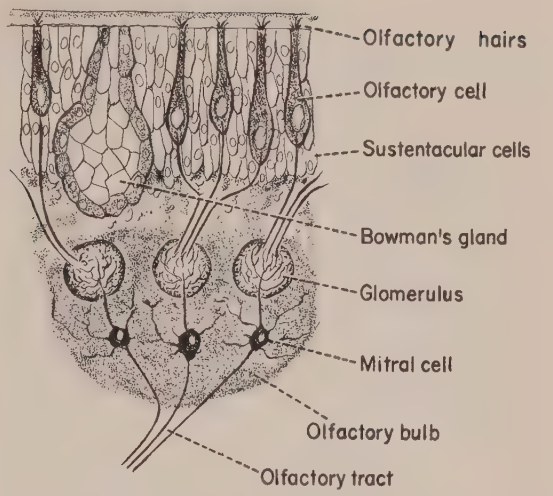


Figure 288. The olfactory membrane, showing especially the olfactory cells with their cilia protruding into the overlying mucus of the nose. (Modified from Bloom and Fawcett: *A Textbook of Histology.* 8th Ed.)

The Olfactory Receptors

The olfactory epithelium contains numerous nerve receptors, called *olfactory cells,* which are shown in Figure 288. These are a special type of nerve cell that projects small root-like structures called *olfactory hairs* outward from the nasal epithelium into the overlying mucus. It is the olfactory hairs that detect the different odors.

The means by which odors excite the olfactory hairs are not well understood. However, those odors most easily smelled are, first, the very highly volatile substances, and, second, substances that are highly soluble in fats. The necessity for volatility is easily understood, for the only means by which an odor can diffuse into the high spaces of the nose is by being carried in the air to this region. The reason for the fat solubility seems to be that the olfactory hairs themselves are outcroppings of the cell membrane of the olfactory cell, and all cell membranes are composed mainly of fatty substances. Presumably, then, the substance becomes dissolved in the olfactory hairs, and this produces nerve impulses in the olfactory cells.

The primary sensations of smell. It has been very difficult to study individual olfactory cells, for which reason we are not yet sure what primary chemical stimuli excite the various types of olfactory cells. Yet on the basis of crude experiments, the following primary sensations have been postulated:

1. The *acetone* group, which includes acetone and amyl and ethyl acetate.
2. The *benzene* group, which includes most of the aromatic hydrocarbons.
3. The *octane* group, which includes the paraffin hydrocarbons and heavy oils.
4. The *dipentane* group, which includes also terpenes such as cedarwood oil and eucalyptus oil.

Thus, smell, like all the other sensations, is probably subserved by a few discrete types of cells that give rise to specific primary olfactory sensations. However, the preceding listing is probably incomplete and perhaps also in error.

Adaptation of smell. Smell, like vision, can adapt tremendously. On first exposure to a very strong odor, the smell may be very strong, but after a minute or more the odor will hardly be noticeable. The olfactory receptors apprise the person of the presence of an odor, but then do not keep belaboring him with its presence. This is especially valuable when one must work in pungent surroundings.

Masking of odors. Unlike the eye's ability to see a number of different colors at the same time, the olfactory system detects the sensation of only a single odor at a time. However, the odor may be a combination of many different odors. If both a putrid odor and a sweet odor are present, the one that dominates the other is the one that has the greater intensity, or, if both are of about equal intensity, the sensation of smell is between that of sweetness and putridness. The ability of a high intensity odor to dominate is called *masking.* This effect is used in hospitals, toilets, and other areas to make pungent surroundings pleasant;

incense may be burned or some odoriferous but pleasant smelling substance may be evaporated into the air to mask the less desirable odors.

Transmission of Smell Sensations into the Central Nervous System

Because smell is a subjective phenomenon that can be studied satisfactorily only in human beings, almost nothing is known about its transmission into the brain. Most of the information available has been gained only in animals by following the course of olfactory impulses into the brain. It has been shown that impulses pass from the olfactory epithelium, first into the pyriform areas of the brain, located bilaterally near the ventral surface of the cerebrum (see Figure 287). From these regions the impulses are then transmitted into the hippocampal gyrus, the cingulate gyrus, the septum, and the prefrontal lobes. It is believed that the *uncus*, part of the pyriform area, might be the primary sensory cortex of smell, and that the other cortical areas are the olfactory association areas.

The impulses that pass into the prefrontal cortex, the hippocampal gyrus, and cingulate gyrus probably provide many of the subjective reactions to odors. For instance, stimulating these areas can frequently cause licking of the lips and varied emotions such as pleasure, docility, and excitement. Therefore, it is probable that the smell signals on reaching the brain divide into several different signals, those to the uncus providing the actual conscious sensations of smell, those to other areas providing reactions such as the unconscious force that draws one to a freshly broiled steak or to an oven-fresh pie.

REFERENCES

Ades, H. W.: Central auditory mechanisms. *Handbook of Physiology* 1:585, 1959.

Adey, W. R.: The sense of smell. *Handbook of Physiology* 1:535, 1959.

Adrian, E. E.: The basis of sensation; some recent studies of olfaction. *Brit. M. J.* 1:287, 1954.

Davis, H.: Excitation of auditory receptors. *Handbook of Physiology* 1:565, 1959.

Davis, H.: Some principles of sensory receptor action. *Physiol. Rev.* 41:391, 1961.

Jackson, C., and Jackson, C. L.: *Diseases of the Nose, Throat, and Ear.* 2nd Ed. Philadelphia, W. B. Saunders Company, 1959.

Kalmus, H., and Hubbard, S. J.: *The Chemical Senses in Health and Disease.* Springfield, Ill., Charles C Thomas, 1960.

Lawrence, M.: Hearing. *Ann. Rev. Physiol.* 23:485, 1961.

Pfaffmann, C.: The sense of taste. *Handbook of Physiology* 1:507, 1959.

Wever, E. G., and Lawrence, M.: *Physiological Acoustics.* Princeton, Princeton University Press, 1954.

Section Eight

THE GASTROINTESTINAL AND METABOLIC SYSTEMS

GASTROINTESTINAL MOVEMENTS AND SECRETION, AND THEIR REGULATION

29

The function of the gastrointestinal system, illustrated in Figure 289, is to provide nutrients for the body. Food, after entering the mouth, is propelled through the esophagus into the stomach, and then through the small and large intestines before finally emptying out the anus. While the food passes through the gastrointestinal tract, digestive enzymes secreted by the gastrointestinal glands act on the food, breaking it into simple chemical substances that can be absorbed through the intestinal wall into the circulating body fluids. The general functions of the gastrointestinal tract, therefore, can be divided into (1) propulsion and mixing of the gastrointestinal contents, (2) secretion of digestive juices, (3) digestion of food, and (4) absorption of food. The first two of these functions are discussed in the present chapter, and the remaining two in the following chapter.

Physiologic Anatomy of the Gastrointestinal Tract

The gastrointestinal tract is essentially a long muscular tube with an inner lining that secretes digestive juices and absorbs nutrients. Figure 290 illustrates a typical cross-section of the gut, show-

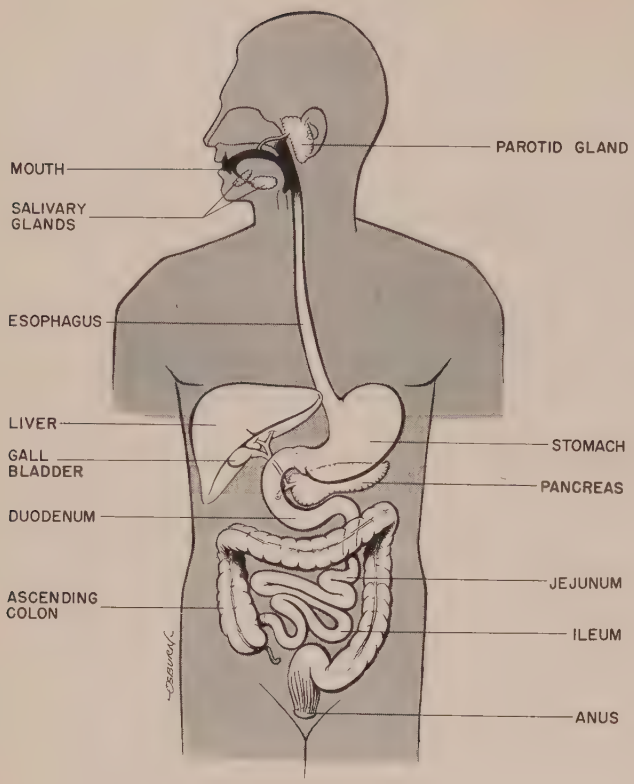


Figure 289. The gastrointestinal tract from the mouth to the anus.

ing that most of the outer part is *smooth muscle* arranged in two layers, a *longitudinal layer* and a *circular layer*. Contraction of the longitudinal muscle shortens the gut, and contraction of the circular layer constricts it. The inner lining of the gut is called the *mucosa*, and it is covered by an *epithelium*. Small glands called *mucosal glands* penetrate into the deeper layers of the lining. These glands secrete digestive juices. Protruding from the lining of parts of the gut are many small *villi* which are responsible for most of the absorption of nutrients.

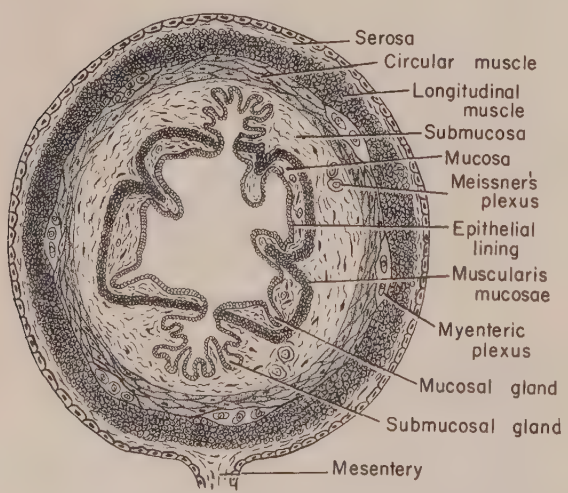


Figure 290. Typical cross-section of the gut.

The intramural nerve plexus. One of the primary controllers of gastrointestinal function is a plexus of nerves called the *intramural nerve plexus* which is divided into two subdivisions, the *myenteric plexus* which lies mainly between the two muscle layers of the gut and *Meissner's plexus* which extends into the inner lining. The intramural plexus is present in the wall of the gut all the way from the esophagus to the anus, forming an intertwining web of nerve fibers and nerve cell bodies. It controls muscular contraction in the gut and controls secretion by many of the glands.

Parasympathetic nerve fibers carried mainly in the vagus nerve terminate in the intramural plexus and, when stimulated, increase the degree of activity of the nerve network. *Sympathetic nerve fibers* also terminate in the intramural plexus and, when stimulated, have exactly the opposite effect on activity of the plexus, decreasing its level of activity.

GASTROINTESTINAL MOVEMENTS

Two basic types of movement occur in the gastrointestinal tract, *propulsive movements* and

mixing movements, one of which keeps the food moving along the gut and the other of which keeps it mixed. The characteristics of the movements in different parts of the gastrointestinal tract exhibit some differences that need to be described separately for certain portions of the gut. However, let us first consider the general characteristics of the movements.

Propulsive Movements of the Gastrointestinal Tract—Peristalsis

Food is moved along the gastrointestinal tract by *peristalsis* which is caused by slow advancement of a circular constriction, as illustrated in Figure 291. This has very much the same effect as encircling one's fingers tightly around a thin tube full of paste and then pulling the fingers along the tube. Any material in front of the fingers will be squeezed forward.

Mechanism of peristalsis. Peristalsis is caused by nerve impulses that move along the myenteric nerve plexus. Stimulation of any single point of the plexus causes impulses to travel around the gut and lengthwise in both directions. The impulses traveling around the gut constrict it, and those traveling lengthwise cause the constriction to move forward. The usual rate of movement of this constriction is a few centimeters per second.

The usual stimulus that initiates peristalsis is *distention* of the gut, for this excites the local myenteric plexus, causing a circular constriction to advance along the gut.

Law of the gut. Even though peristalsis can move in both directions along the gut, it most frequently moves toward the anus. The probable reason for this is that the myenteric plexus itself is "polarized" in this direction. When a portion of the gut becomes distended, it causes contraction of the gut on the headward side of the distention and relaxation on the anal side. The contraction pushes the food forward and the relaxation allows easy forward movement; the movement, therefore, is usually analward rather than backward. This is called the *law of the gut.*

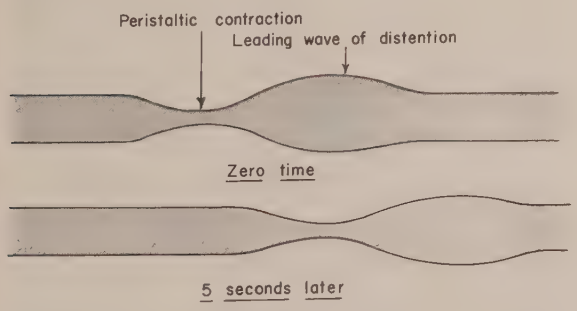


Figure 291. Peristalsis.

Mixing Movements in the Gastrointestinal Tract

The mixing movements consist of two basic types, (1) *weak peristaltic movements* that fail to move the food forward but nevertheless do succeed in mixing the intestinal contents adjacent to the wall of the gut, and (2) *segmental movements,* which are isolated constrictions that occur at many points along the gut at the same time. The precise characteristics of the mixing movements will be described specifically for each part of the gastrointestinal tract.

Let us now begin at the upper end of the gastrointestinal tract and describe both the propulsive and mixing movements as the food passes analward.

Swallowing

Swallowing is initiated by a bolus of food pushed backward on the tongue into the pharynx. The bolus stimulates *swallowing receptor areas* located all around the opening of the pharynx, and impulses pass to the brain stem to initiate a series of automatic muscular contractions as follows:

(1) The soft palate is pulled upward to close the posterior part of the nose from the mouth.
(2) The vocal cords and larynx close strongly, and the epiglottis swings backward to prevent food from going into the trachea.
(3) The larynx is pulled upward, and at the same time the muscular sphincter that normally keeps the esophagus closed is relaxed. These two effects open the esophagus.
(4) Constriction of the pharyngeal muscles forces the bolus from the pharynx downward into the esophagus.

NERVOUS CONTROL OF SWALLOWING. Figure 292 illustrates the nervous pathways involved in the swallowing mechanism, showing that the swallowing receptors transmit impulses mainly through the *trigeminal nerve* into the reticular substance of the *medulla oblongata* where the swallowing center is located. Once this center has been activated, the sequence of muscular reactions listed above occurs automatically and usually cannot be stopped after it has been started. Nerve impulses go by way of the *glossopharyngeal* and *vagus nerves* to the pharyngeal

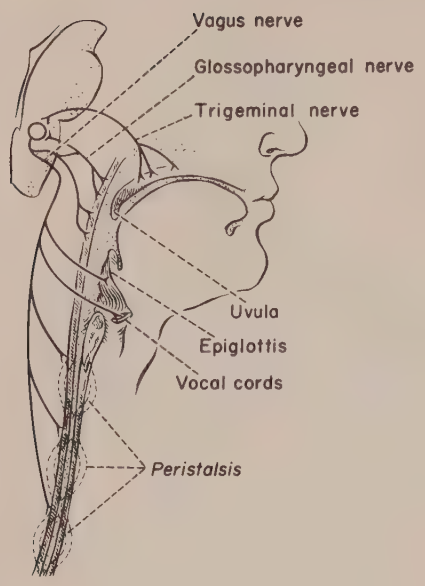


Figure 292. Anatomy of the swallowing mechanism.

and laryngeal regions to move the bolus of food into the upper esophagus. Then impulses from the vagi activate the proximal portion of the esophagus to push the food on toward the stomach.

Esophageal stage of swallowing. The musculature of the pharynx and of the upper one third of the esophagus is different from that of the rest of the gastrointestinal tract, for this muscle is skeletal muscle, controlled directly by nerves from the brain in contrast to the remainder of the esophagus and gastrointestinal tract that is composed of smooth muscle and is only indirectly controlled by the central nervous system through the effects of the autonomic nervous system on the intramural plexus. Therefore, all contractions of the pharynx and upper one third of the esophagus are initiated directly by vagal and glossopharyngeal nerve impulses, and without these nerves the act of swallowing becomes paralyzed.

However, once food has reached the middle third of the esophagus, the distention elicits a typical peristaltic wave called the *secondary peristaltic wave* of the esophagus that pushes the food the rest of the way into the stomach. The entire time required for food to pass from the pharynx to the stomach is about five to ten seconds.

Function of the cardiac constrictor. The centimeter or more of the esophagus that is proximal to the stomach is normally maintained in a state of tonic contraction. This area of the esophagus is called the *cardiac constrictor* or simply the *cardia;* it is illustrated in Figure 293.

When swallowed food reaches the lower end of the esophagus, its progress is momentarily slowed by the cardia, but the approach of the esophageal peristaltic wave causes relaxation of the cardia which allows the food to move on into the stomach.

On the other hand, contractions in the stomach ordinarily will not push food backward through the cardia because there normally is no backward relaxation effect. Fortunately, therefore, stomach contractions normally propel food only forward rather than backward.

Motor Functions of the Stomach

The motor functions of the stomach are threefold: (1) storage of large quantities of food immediately after a meal, (2) mixing of this food with gastric secretions, and (3) emptying of the food from the stomach into the small intestine. The basic functional parts of the stomach are illustrated in Figure 293. Physiologically the stomach can be divided into two major parts, the *corpus* and the *antrum.*

Storage function of the stomach. Food, on emptying from the esophagus into the stomach, first enters the corpus which is an extremely elastic bag that can store very large quantities of food. Furthermore, the tone of the corpus is normally very slight, so that even extreme amounts of food do not increase the pressure in the stomach greatly. Thus, the corpus is mainly a receptive organ for holding food until it can be utilized by the remainder of the gastrointestinal tract.

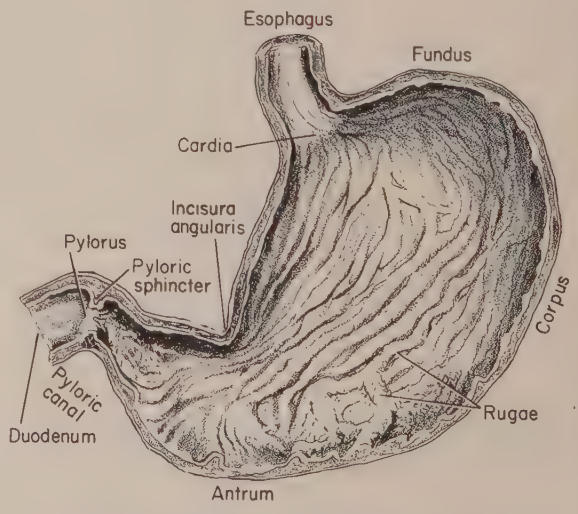


Figure 293. Physiological anatomy of the stomach.

Mixing in the stomach—chyme. The gastric glands, which cover most of the mucosa of the corpus, secrete large quantities of digestive juices which come into contact immediately with the stored food. Weak, rippling peristaltic waves, called *tonus waves* or *mixing waves*, pass along the stomach wall approximately once every 20 seconds. These begin in any part of the corpus and spread for variable distances toward the antrum, and they become more intense when food is present in the stomach than when it is not present.

The mixing waves mix the gastric secretions with the outermost layer of food and gradually move the mixture toward the antral part of the stomach. On entering the antrum, the waves become stronger, and the food and gastric secretions become mixed to a greater and greater degree of fluidity. As the food becomes thoroughly mixed with the gastric secretions, the mixture takes on a milky white appearance and is then called *chyme*.

Propulsion of chyme through the stomach and emptying of the stomach. The tonus or mixing waves are rarely strong enough to push chyme through the pylorus into the duodenum. However, occasional very powerful peristaltic contractions also occur, beginning either in the corpus or antrum, and these generate as much as 50 mm. Hg pressure in the prepyloric portion of the antrum. This is enough pressure to push open the pyloric sphincter and propel the chyme on into the duodenum.

REGULATION OF STOMACH EMPTYING. Emptying of the stomach is controlled mainly by the intensity of the strong peristaltic waves. The pylorus itself usually remains slightly constricted all the time. Therefore, a certain degree of pressure is required before chyme can pass through. A weak peristaltic wave fails to move the chyme but a strong peristaltic wave succeeds. Among the different factors that determine whether or not the peristaltic wave will succeed are the following:

Degree of fluidity of the chyme. Obviously, the better the food has become mixed with gastric secretions the more easily it can flow through the narrow passageway of the pylorus. Therefore, ordinarily, food will not pass out of the stomach until it has been thoroughly mixed.

Quantity of chyme already present in the small intestine. When a large amount of chyme has already emptied into the small intestine, particularly when a large portion of this is still present in the duodenum, a reflex called an *enterogastric reflex* spreads backward through the myenteric plexus from the duodenum to the stomach to inhibit peristalsis. In this way, the duodenum keeps itself from becoming overfilled.

Presence of acids and irritants in the small intestine. The gastric secretions, as will be discussed later in the chapter, are highly acidic, but the acid in the chyme entering the duodenum is ordinarily neutralized by pancreatic secretions that also empty into the duodenum. Until the acid becomes neutralized by pancreatic juice, irritation of the wall of the duodenum elicits an enterogastric reflex similar to that elicited by distention; this too inhibits the peristaltic waves in the stomach, thus stopping gastric emptying. In this way, the duodenum protects itself from too much acidity. Likewise, any other irritant also causes an enterogastric reflex that will do the same thing.

Presence of fats in the small intestine—enterogastrone. When fats enter the small intestine from the stomach, they extract from the mucosa of the duodenum and jejunum a hormone called *enterogastrone*, which is immediately absorbed into the blood and carried to the stomach. Here it inhibits stomach peristalsis and slows up stomach emptying. This mechanism allows adequate time for fat digestion to occur in the small intestine. Proteins and carbohydrates, on the other hand, have very little inhibitory effect on stomach emptying. Fortunately, both of these foods are digested much more easily in the intestine than is fat.

In summary, gastric emptying is determined mainly by fluidity of the contents in the stomach and the state of the duodenum. If the duodenum is already filled, has irritant substances in it, or has fat in it, emptying will proceed very slowly, but if the duodenum is empty and the contents of the stomach are very fluid, emptying will proceed rapidly.

Hunger contractions. In addition to the tonus and peristaltic contractions of the stomach, a third type of very intense contraction, called *hunger contractions*, occur when the stomach has been empty for 8 to 20 hours. These are normally very powerful rhythmic peristaltic contractions but occasionally become a strong tetanic contraction lasting for as long as 2 to 10 minutes. They cause a tight sensation in the pit of the stomach called *hunger pangs*, and they intensify the desire for food.

Movements of the Small Intestine

Mixing contractions—segmentation. The presence of chyme in the small intestine initiates a type of contraction called *segmentation*, which is illustrated in Figure 294. When the small

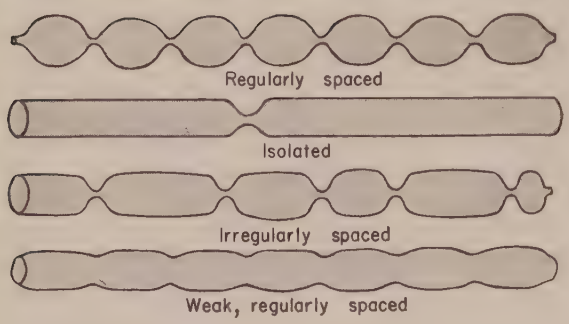


Figure 294. Segmentation movements of the small intestine.

intestine becomes distended, many constrictions occur either regularly or irregularly along the distended area. As shown in Figure 294, the intestine becomes "chopped" into small sausage-like vesicles. The constrictions then relax, but others occur at different points a few seconds later. Thus, repetitive "chopping" of the chyme keeps it mixed continually while it is in the small intestine.

Propulsive movements. It is in the small intestine that the most typical peristalsis occurs, for distention of any portion of the small intestine with chyme initiates a peristaltic wave. Peristalsis is far more intense when the parasympathetic nerves are stimulated, and sympathetic stimulation can inhibit greatly or totally block peristalsis.

Both the segmentation and peristaltic movements of the small intestine are controlled by the myenteric plexus.

Emptying of intestinal contents at the ileocecal valve. Figure 295 shows the *ileocecal valve* where the small intestine empties into the large intestine. Note that the small intestine actually protrudes forward into the colon. This projection of the ileocecal valve prevents the contents of the colon from regurgitating into the small intestine.

Emptying of the small intestine at the ileocecal valve occurs in very much the same way that the stomach empties; that is, peristaltic waves in the small intestine build up pressure behind the valve and push chyme forward into the colon, but if the colon has become too full, myenteric reflexes can inhibit peristalsis and thereby slow up or stop the emptying.

Movements of the Colon

The functions of the colon are (1) absorption of water and electrolytes from chyme and (2) storage of fecal matter until it can be expelled. The half of the colon proximal to the small intestine, which is illustrated in Figure 289, is concerned mainly with absorption and the distal half with storage. Therefore, except when the bowels are to be emptied, the movements of the colon are usually very sluggish.

Mixing movements. The mixing movements of the colon are similar to the segmentation movements of the small intestine. Concentric contractions of the colon break it up into large pockets called *haustrations,* which are illustrated in Figures 289, 295, and 296. The circular constrictions last for about 30 seconds, and after another few minutes occur in nearby but not the same areas. Thus, fecal material is slowly "dug" into and rolled over in much the same manner that one spades the earth.

Ordinarily, 450 milliliters of chyme is emptied into the colon each day, and about 370 milliliters of this, mainly the water and electrolytes, is reabsorbed before defecation takes place.

Propulsive movements. The typical peristaltic movements, as in the small intestine, do not occur in the colon, for the colon has no peristaltic movements at all 99 per cent of the time. Yet when the colon becomes overfilled, several strong peristaltic movements, called *mass movements,* occur one after another. These propel the fecal material long distances, sometimes all the way from the ascending colon to the descending colon. After a few minutes the mass movements cease, to reappear again many hours later when some part of the colon becomes overfilled again.

Defecation. When the mass movements of the colon have succeeded in moving fecal material into the rectum, a special reflex called the *defeca-*

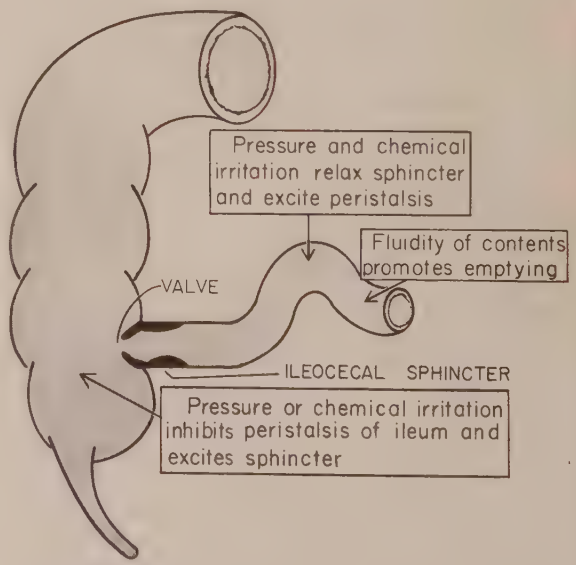


Figure 295. Emptying at the ileocecal valve.

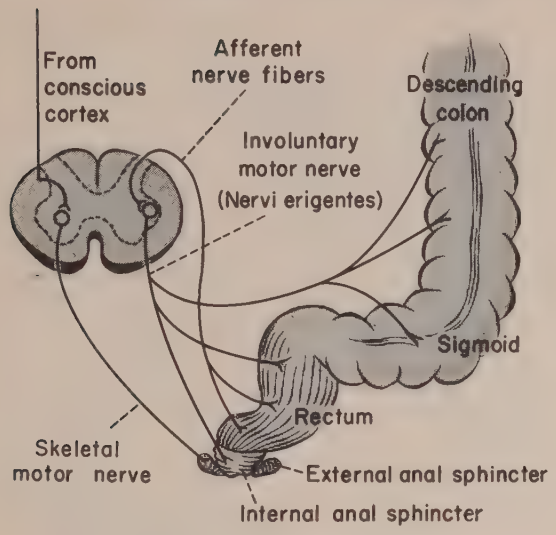


Figure 296. Anatomy of the defecation reflex.

tion reflex occurs, which causes emptying of the rectum and lower parts of the colon. Figure 296 shows that filling the rectum excites nerve endings that send impulses into the lower part of the spinal cord. These cause reflex impulses to be transmitted through the sacral parasympathetic nerves to the descending colon, sigmoid flexure, rectum, and *internal anal sphincter,* causing relaxation of the sphincter and contraction of the gut wall. This reflex thus causes emptying of the bowels if the *external anal sphincter* is also relaxed. However, the external anal sphincter is a skeletal muscle controlled by voluntary skeletal nerves that can be relaxed or tightened according to the will of the person. If the time is not propitious for emptying the bowels, tightening of the external anal sphincter prevents defecation despite the defecation reflex. On the other hand, relaxation allows defecation to take place.

If a person prevents defecation when the defecation reflex occurs, the reflex usually dies out after a few minutes but returns a few hours later. Also, a person can frequently initiate a defecation reflex at will by tightening his abdominal muscles, which compresses the rectal wall and elicits the typical reflex. Unfortunately, this elicited reflex is usually much weaker than the natural reflex, so that defecation is less efficacious under these conditions than when the reflex has been elicited naturally.

Special Types of Gastrointestinal Movements

Antiperistalsis and vomiting. Occasionally, some intensely irritating substance enters the gastrointestinal tract. An immediate effect of this is an increase in the rate of local secretion of mucus, which helps to protect the interior of the gut. Simultaneously, the local gut wall contracts intensely. For reasons not too well understood, these intense local contractions elicit *antiperistalsis,* meaning peristalsis backward toward the mouth, rather than forward peristalsis. Food cannot move backward from the colon into the small intestine because of the ileocecal valve, but it can move all the way from the tip of the small intestine back into the stomach.

On reaching the stomach, the irritative material is then rapidly expelled by the vomiting process as follows: The intense irritation of the gut causes impulses to be transmitted through the visceral sensory nerves into the brain. These cause a sensation of *nausea,* and, if the impulses are strong enough, they will also cause an automatic reflex integrated in the medulla oblongata called the *vomiting reflex.* This reflex causes closure of the airway into the trachea, relaxation of the cardia of the stomach, and very tight contraction of both the diaphragm and the abdominal muscles. The squeezing action of the diaphragm and abdominal muscles on the stomach pushes food out of the stomach upward through the esophagus and mouth. This is the vomiting process.

Gastrocolic and duodenocolic reflexes. Almost everyone is familiar with the natural desire to defecate following either a heavy meal or the first meal of the day. The cause of this is the *gastrocolic* and *duodenocolic reflexes,* mainly the latter. These reflexes are elicited by increased filling of the stomach and duodenum, which transmits impulses downward along the myenteric plexus to the colon to cause increased excitability of the entire colon, initiating both mass movements and defecation reflexes.

The peritoneal reflex. Irritation of the peritoneum in any way, whether caused by cutting of the peritoneum in an abdominal operation, by infection in the peritoneum, or even by a severe blow to the abdomen that causes trauma of the peritoneum, will elicit a *peritoneal reflex* that strongly excites the sympathetic nerves to the gut. These nerves in turn *inhibit* gastrointestinal activity and thereby stop all movement of chyme along the intestinal tract. Obviously, lack of movement aids in the repair of the peritoneal damage.

Mucosal reflexes. Irritation inside the gut or distension of the gut ordinarily excites the intramural plexus rather than inhibiting it. Reflexes occur mainly locally, causing increased local secretion and increased local motor activity. The increased secretion dilutes the irritating factor, and the motor activity moves the irritant on through the gut. When the irritation is great, it causes *diarrhea.* And if the degree of irritation

increases still more, it can cause diarrhea from the lower part of the intestinal tract while causing vomiting from the upper intestinal tract. Obviously, these reflexes are protective in nature and prevent damage from irritating substances or infectious processes inside the intestinal tract.

GASTROINTESTINAL SECRETIONS

Glands are present throughout the gastrointestinal tract to secrete chemicals that mix with the food and digest it. These secretions are of two types: first, mucus which protects the wall of the gastrointestinal tract, and, second, enzymes and allied substances that break the large chemical compounds of the food into simple compounds.

Mucus. Mucus is secreted by every portion of the gastrointestinal tract. It contains a large amount of mucoprotein that is resistant to almost all digestive juices. Mucus lubricates the passage of food along the mucosa, and it forms a thin film everywhere to prevent the food from excoriating the mucosa. Also, it is *amphoteric*, which means it is capable of neutralizing either acids or bases. All these properties of mucus make it an excellent substance to protect the mucosa from physical damage, and to prevent the digestion of the wall of the gut by the digestive juices.

Salivary Secretion

Saliva is secreted by the *parotid, submaxillary, sublingual,* and other smaller glands in the mouth. Saliva is about half *mucus* and half a solution of the enzyme *ptyalin*. The function of mucus is to provide lubrication for swallowing. Without mucus one can hardly swallow. If one mixes food with water to take the place of mucus, approximately 10 times as much water as mucus is necessary to provide the same degree of lubrication. The function of the ptyalin in the saliva is to begin the digestion of starches and other carbohydrates in the food. Ordinarily the food is not exposed in the mouth to saliva long enough for more than 5 to 10 per cent of the starches to become digested. However, the mixed saliva and food is normally stored in the corpus of the stomach for about 30 minutes to an hour, during which time the saliva may digest as much as 40 to 50 per cent of the starches.

Regulation of salivary secretion. The *superior and inferior salivatory nuclei* located in the brain stem, as shown in Figure 297, control secretion by the salivary glands. These nuclei in turn are

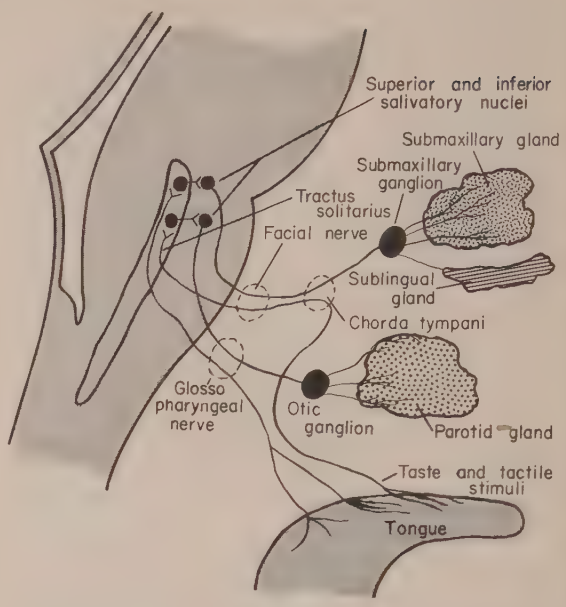


Figure 297. Nervous regulation of salivary secretion.

controlled mainly by taste impulses and other sensory impulses from the mouth. Foods that have a pleasant taste ordinarily cause the formation of large quantities of saliva, while some unpleasant foods may decrease salivary secretion so greatly that swallowing is made very difficult. Also, the sensation of smooth-textured foods inside the mouth increases salivation, while the sensation of roughness decreases salivation. This effect presumably allows those foods that will not abrade the mucosa to be swallowed with ease, and causes the rejection of abrasive foods.

The phases of salivary secretion. In addition to the salivation that occurs while food is actually in the mouth, salivation frequently occurs even before food enters the mouth—that is, when a person is thinking about or smelling pleasant food—and it continues to occur even after the food has been swallowed. Therefore, salivary secretion can be divided into three phases, the *psychic phase,* the *gustatory phase,* and the *gastrointestinal phase.* The psychic phase presumably makes the mouth ready for food and aids in the secretion of saliva as the food is presented to the mouth. The gustatory phase supplies the saliva that mixes with the food while one is chewing, and the gastrointestinal phase continues the secretion of saliva even after the food has passed for storage into the stomach. Secretion during the gastrointestinal phase is especially likely to be abundant when one has swallowed irritant foods. The saliva, on being swallowed, helps to neutralize the irritant substance, thereby relieving any irritation of the stomach that might be occurring.

Esophageal Secretions

The esophagus secretes only *mucus*. Normally, food passes from the mouth through the esophagus and into the stomach in about 7 seconds. This food has not been subjected to the mixing movements of the gastrointestinal tract and, therefore, is in its most abrasive state. Fortunately the esophagus is supplied with a great abundance of mucus glands that secrete mucus to protect the mucosa from excoriation.

Gastric Secretions

Mucus. The primary function of the gastric secretions is to begin the digestion of proteins. Unfortunately, the wall of the stomach is composed of protein substances. Therefore, the surface of the stomach must be exceptionally well protected at all times against digestion. Perhaps the most abundant secretion of mucus in any part of the gastrointestinal tract occurs in the stomach. The entire surface of the stomach is covered by very small *mucous cells*, the outer half of which are composed almost entirely of mucus; these prevent gastric secretions from ever touching the deeper layers of the stomach wall. In the antral region of the stomach, where the powerful peristaltic movements occur and where excoriation of the stomach wall is particularly likely to occur, mucus is secreted not only by the mucous cells on the surface of the mucosa but also by large mucous glands that extend deep into the mucosa. In the absence of mucus secretion, a hole is eaten in the wall of the stomach in only a few hours. This hole is called a *stomach ulcer*.

Digestive substances. The major digestive substances secreted by the stomach are *hydrochloric acid* and *pepsin*. Hydrochloric acid activates the pepsin, which is an enzyme that begins the digestion of proteins.

Less abundant enzymes secreted by the stomach are *gastric lipase* for beginning the digestion of fats, and *rennin* for aiding in the digestion of *casein*, one of the proteins in milk. These enzymes are secreted in such minor quantities that they are of almost no importance.

The total quantity of stomach secretion each day is about 2000 ml.

Regulation of gastric secretion. NEUROGENIC MECHANISMS. Stomach secretion is regulated by both neurogenic and hormonal mechanisms, shown in Figure 298. Some of the neurogenic mechanisms are quite similar to those regulating salivary secretion. They employ reflexes from the medulla that transmit secretion signals through the vagus nerves to the stomach, and they also employ local myenteric reflexes. Both of these types of reflexes can be initiated by food stimulating the mucosa of the stomach, but the medullary reflexes can also be excited by impulses

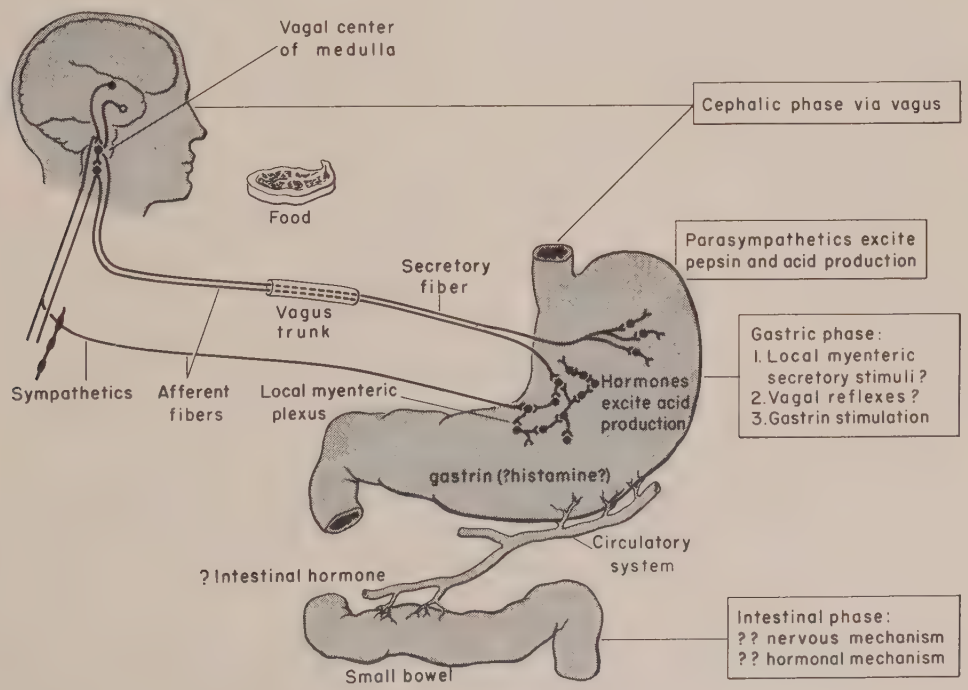


Figure 298. Regulation of gastric secretion.

from various areas of the brain, particularly the cerebral cortex.

THE GASTRIN MECHANISM. Gastric secretion is also regulated to some extent by a hormone called *gastrin.* When meats and certain other protein foods reach the antral portion of the stomach, they cause the hormone *gastrin* to be extracted from the antral mucosa and to be absorbed into the blood stream. This hormone, which is very similar to the substance histamine, then passes by way of the blood to the fundic glands of the stomach, and causes them to secrete a strongly acidic gastric juice. The acid, in turn, greatly aids in the digestion of the meats that first initiated the gastric mechanism. In this way the stomach helps to tailor-make the secretion to fit the particular type of food that is eaten.

THE PHASES OF GASTRIC SECRETION. Large amounts of stomach juices are often secreted when pleasant food is simply thought of or particularly when it is smelled. This is called the *psychic phase* of stomach secretion. It prepares the stomach for food that is to be eaten. The second phase of gastric secretion is the *gastric phase,* which is the secretion that occurs while the food is in the stomach itself. This is caused by reflexes initiated by food in the stomach and by the gastrin mechanism. Finally, even after the food has left the stomach, gastric secretions continue for several hours. This is called the *intestinal phase* of gastric secretion. It is probably caused by hormones that pass to the stomach in the blood after being extracted from the intestinal mucosa by the food. Ordinarily, the amount of secretion during the intestinal phase is only about 10 per cent of the total during the other two phases.

Pancreatic Secretions

The pancreas, shown in Figure 299, is a large gland located immediately beneath the stomach. It empties about 1200 ml. of secretions each day into the upper portion of the small intestine 1½ inches beyond the pylorus. These secretions contain large quantities of *amylase* for digesting carbohydrates, *trypsin* and *chymotrypsin* for digesting proteins, *pancreatic lipase* for digesting fats, and other less important enzymes. It is obvious from this list that the pancreatic secretions are as important for digesting the food as any others of the entire gastrointestinal tract.

In addition to the digestive enzymes, pancreatic secretions contain large amounts of *sodium bicarbonate,* which reacts with the hydrochloric acid from the stomach to form sodium chloride and carbonic acid. The carbonic acid then is absorbed into the blood, becomes water and carbon dioxide, and the carbon dioxide is expired through the lungs. The net result is an increase in the quantity of sodium chloride, a neutral salt, in the intestine. Thus, pancreatic secretions neutralize the acidity of the chyme coming from the stomach. This is one of the most important functions of the pancreas.

Regulation of pancreatic secretion. THE SECRETIN MECHANISM AND NEUTRALIZATION OF CHYME. When chyme enters the upper small intestine it causes a polypeptide called *secretin* to be released from the intestinal mucosa; the quantity of secretin released is especially abundant when the chyme is highly acidic. The secretin in turn is absorbed into the blood and carried to the glandular cells of the pancreas. There it causes

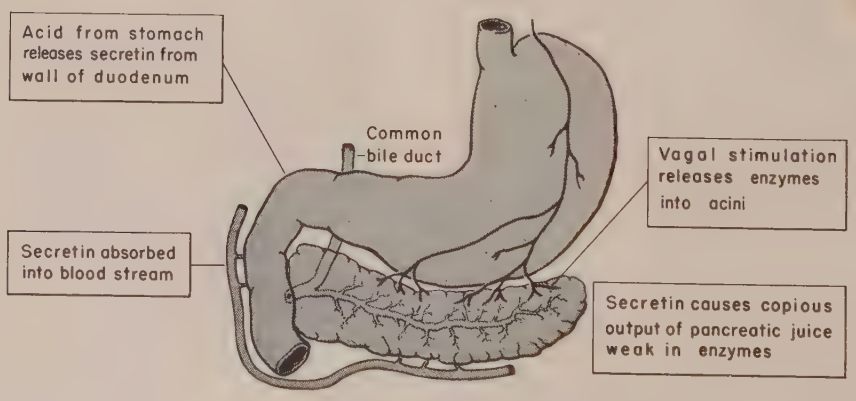


Figure 299. Regulation of pancreatic secretion.

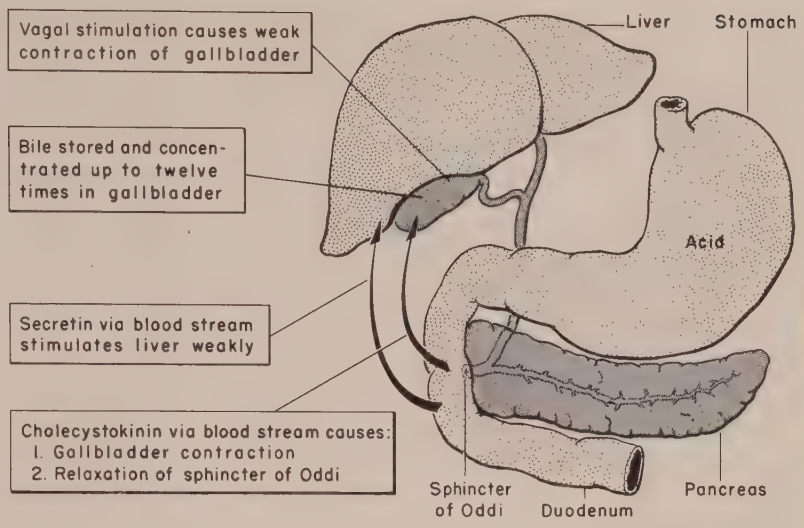


Figure 300. Bile secretion, bile storage in the gallbladder, and the cholecystokinin mechanism for promoting gallbladder emptying.

the cells to secrete large quantities of fluids containing extra large amounts of sodium bicarbonate. The secretin mechanism, therefore, causes automatic neutralization of the chyme in the upper small intestine.

When satisfactory neutralization does not occur, the acidic chyme, containing also large quantities of pepsin, is likely to erode through the wall of the duodenum, the uppermost part of the small intestine, causing an ulcer. In fact, ulcer of the duodenum is about four times as common as ulcer of the stomach, because this area is not nearly as well protected by mucous glands as is the stomach.

THE PANCREOZYMIN MECHANISM. At the same time that secretin is extracted from the intestinal mucosa, another hormone, *pancreozymin,* also is extracted in response mainly to proteins in the chyme, but to a less extent in response to fats and carbohydrates. Pancreozymin, like secretin, passes by way of the blood to the pancreas, but, unlike secretin, it causes the secretory cells to secrete large quantities of digestive enzymes instead of sodium bicarbonate. These enzymes, on entering the duodenum, begin digesting the foods.

VAGAL REGULATION. Stimulation of the vagus nerve also causes the secretory cells of the pancreas to secrete highly concentrated enzymes. The quantity, however, is usually so small that the enzymes remain in the ducts of the pancreas and later are floated into the intestinal tract by the copious secretion of fluid that follows secretin stimulation.

Vagal stimulation of pancreatic secretion seems to be a by-product of the vagal reflexes to the stomach. That is, some of the reflex impulses initiated by food in the stomach return to the pancreas rather than to the stomach. This allows preliminary formation of pancreatic enzymes even before the food enters the intestine. However, vagal stimulation of pancreatic secretion is probably unimportant in comparison with the hormonal stimulation by secretin and pancreozymin.

Liver Secretion

The liver, shown in Figure 300, secretes a solution called *bile* that contains a large quantity of *bile salts,* a moderate quantity of *cholesterol,* a small quantity of the green pigment *bilirubin,* which is a waste product of red blood cell destruction, and a number of other less important substances. The only substance in bile that is of importance to the digestive functions of the gastrointestinal tract is the bile salts. The remainder of the contents is actually waste products being removed from the body fluids by this route of *excretion.*

The bile salts are not enzymes for digesting foods, but they act as a powerful *detergent* (a substance that lowers the surface tension of the surface between water and fats). This helps the mixing movements of the intestine break the large fat globules of the food into small globules, thus allowing the lipases of the intestinal tract, which are water soluble, to attack larger surface areas of the fat, and to digest it. Without this action of bile almost none of the fats in the food would be digested.

Regulation of bile secretion. Liver secretion, unlike secretion by other gastrointestinal glands, does not increase and decrease significantly in response to food in the intestine. The only hor-

mone known to affect it is secretin, which can increase the output of bile some 10 to 20 per cent, but does not have the powerful effect on the liver that it has on the pancreas.

STORAGE OF BILE IN THE GALLBLADDER. Even though bile secretion is a continuous process, the flow of bile into the gastrointestinal tract is not continuous. Figure 300 illustrates the *sphincter of Oddi* that normally blocks the flow of bile from the *common bile duct* into the duodenum. Instead of flowing into the gut, the bile flows into the *gallbladder,* which is attached to the common bile duct. Much of the fluid and electrolytes of the bile is then reabsorbed into the blood by the gallbladder mucosa. This concentrates as much as twelve-fold the bile acids, cholesterol, and bilirubin, which cannot be absorbed, and allows the gallbladder, even though it has a maximum volume of only 50 ml., to accommodate the active components (bile acids) of an entire day's liver secretion of bile (600 ml.).

EMPTYING OF THE GALLBLADDER. When food enters the small intestine, two mechanisms simultaneously cause the gallbladder to empty into the small intestine. First, the fatty substances in the food extract from the intestinal mucosa a hormone called *cholecystokinin,* which then passes through the blood to the gallbladder and causes the muscular wall to contract. Second, the presence of food in the duodenum causes peristalsis in this region, and the peristaltic waves periodically open the sphincter of Oddi. The combination of gallbladder contraction and opening of the sphincter of Oddi allows the stored bile to empty into the intestine, and the bile salts immediately begin their emulsifying action on the fats.

GALLSTONES. Gallstones are caused mainly by a fatty waste product, *cholesterol,* which is normally excreted in the bile. Cholesterol is relatively insoluble, but is normally held in solution in the bile by chemical attraction to the bile salts. Often, though, when bile becomes concentrated in the gallbladder, too much water is removed from the bile, and the cholesterol becomes too concentrated to remain in solution. Small crystals of cholesterol then begin to precipitate and sometimes grow to fill the entire bladder.

A means for preventing the formation of gallstones is to eat a diet low in fat, for cholesterol is formed in great abundance with a high fat diet. Once gallstones have been formed, however, the only treatment is removal of the stones, or, preferably, removal of the gallbladder itself along with the stones. Absence of the gallbladder does not greatly affect the digestion of fats, because bile continues to be excreted into the intestine, though now it enters the gut almost all the time rather than periodically.

Secretions of the Small Intestine

The glands of the small intestine secrete the enzymes *sucrase, maltase,* and *lactase* for splitting disaccharides into monosaccharides, the final digestion products of carbohydrates. Also, these glands secrete large quantities of *peptidases* for performing the final steps in protein digestion, and small quantities of *lipases* for splitting fats.

In addition, the small intestine secretes along its entire surface large quantities of mucus, which provide the same protective function in this part of the gastrointestinal tract as in the stomach, the esophagus, and elsewhere. In the duodenum, an especially abundant amount of mucus is secreted by large mucous glands, called *Brunner's glands,* lying deep in the mucosa. The function of this secretion is to protect this portion of the intestinal tract from the powerful digestive action of pepsin and hydrochloric acid in the chyme newly arrived from the stomach. Once the chyme has been neutralized by pancreatic juice, however, it no longer has such a strong tendency to digest the wall of the intestine, which explains why Brunner's glands are needed only in this uppermost region of the intestinal tract.

Quantity of secretion. The total amount of secretion of the small intestine is about 3 liters per day, which compares with about 500 ml. of saliva, 2000 ml. of gastric juices, 1200 ml. of pancreatic juices, and 600 ml. of bile. In other words, the total quantity of intestinal secretion is almost as much as that of all the remaining gastrointestinal glands put together.

Absorption from the small intestine. Almost all the secretions that enter the small intestine, including saliva, gastric secretions, pancreatic secretions, bile, and intestinal secretions, are absorbed before entering the large intestine. Any fluid that is ingested is absorbed as well. That is, 8 liters or more of fluid is absorbed by the mucosa of the small intestine each day. A remaining 450 ml. passes along with the chyme into the large intestine, and most of it is absorbed there before the feces are expelled. Thus, fluid circulation occurs continually between the body fluids and the gastrointestinal tract. Because the fluid utilized in forming the gastrointestinal secretions is mainly extracellular fluid, any loss of fluids from the gastrointestinal tract, such as by vomiting or diarrhea, is actually a loss of extracellular fluids and can cause one to become extremely dehydrated.

Regulation of secretion of the small intestine. Secretion by each portion of the small intestine is regulated partly by local nervous reflexes and partly by hormones extracted from the intestinal mucosa. The local nervous reflexes are probably responsible for most of the secretion. That is, food distending the intestinal tract or irritating the intestinal mucosa initiates reflexes in the intramural plexus to stimulate secretion by the intestinal glands.

The hormonal mechanism for regulating intestinal secretion, though perhaps not as important quantitatively as the nervous reflex method, helps to determine the types of enzymes secreted. Food in the small intestine extracts a mixture of hormones called *enterocrinin* from the mucosa, and these hormones then stimulate the intestinal glands to secrete appropriate enzymes for digesting the types of food present. For instance, a large quantity of proteins in the intestine causes the concentration of peptidases in the intestinal juices to be greater than the concentration of the other enzymes. Likewise, carbohydrates promote the secretion of sucrase, maltase, and lactase, and fats promote the secretion of lipases.

Secretions of the Large Intestine

The large intestine, like the esophagus, performs no digestive functions. Therefore, its only secretion is mucus. The entire mucosa is coated with mucus cells that provide lubrication for the passage of feces from the ileocecal valve to the anus and also protect the large intestine from digestion by the enzymes emptied from the small intestine. The portions of the large intestine proximal to the ileocecal valve are usually protected against the digestive hormones better than the distal portions. Consequently, during severe diarrhea the rapid flow of enzymes of the small intestine into the distal colon is very likely to cause extreme irritation. Prolonged and severe diarrhea sometimes initiates a condition called *ulcerative colitis,* which occasionally leads to holes in the colon that cause death. The mucus secreted in the large intestine normally protects against this.

REFERENCES

Babkin, B. P.: *Secretory Mechanism of the Digestive Glands.* 2nd Ed. New York, Paul B. Hoeber, Inc., 1950.

Code, C. F., et al.: *An Atlas of Esophageal Motility in Health and Disease.* Springfield Ill., Charles C Thomas, 1958.

Grossman, M. I.: The digestive system. *Ann. Rev. Physiol.* 25:165, 1963.

Grossman, M. I.: The glands of Brunner. *Physiol. Rev.* 38:675, 1958.

Hunt, J. N.: Gastric emptying and the secretion in man. *Physiol. Rev.* 39:491, 1959.

Jones, F. A.: *Modern Trends in Gastroenterology.* New York, Paul B. Hoeber, Inc., 1958.

Lundberg, A.: Electrophysiology of salivary glands. *Physiol. Rev.* 38:21, 1958.

Popper, H., and Schaffner, F.: *Liver: Structure and Function.* New York, McGraw-Hill Book Company, Inc., 1957.

Taylor, W. H.: Proteinases of the stomach in health and disease. *Physiol. Rev.* 42:519, 1962.

Terracol, J., and Sweet, R. H.: *Diseases of the Esophagus.* Philadelphia, W. B. Saunders Company, 1958.

Woodward, E. R., and Dragstedt, L. R.: Role of the pyloric antrum in regulation of gastric secretion. *Physiol. Rev.* 40:490, 1960.

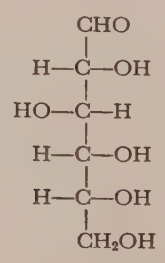
DIGESTION AND ASSIMILATION OF CARBOHYDRATES, FATS, AND PROTEINS

30

The term *digestion* means the splitting of large chemical compounds in the foods into simpler compounds that can be used by the body. The term *assimilation* includes several functions that may be listed: (1) absorption of the digestive end-products into the body fluids, (2) transport of these to the cells where they will be used, and (3) chemical change of some of them into other substances that are specially needed for various purposes. The function of the digestive and assimilative processes is to provide nutrients for the chemical reactions of metabolism.

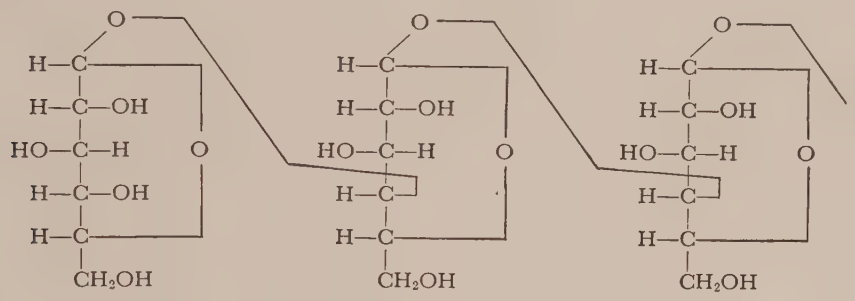
DIGESTION, ABSORPTION, AND DISTRIBUTION OF CARBOHYDRATES

Carbohydrates are composed of carbon, hydrogen, and oxygen. The basic structure of a carbohydrate is a *monosaccharide,* the most common of which is *glucose,* which has the following chemical formula:



Two other monosaccharides very frequently present in food are *fructose* and *galactose,* the formulas for which are the same as that of glucose except that some of the "H" and "OH" radicals are transposed.

Glucose and other monosaccharides are usually *polymerized* (combined together) into larger chemical compounds such as *starches, glycogens, pectins,* and *dextrins.* By far the most common carbohydrate of the diet is starch, which is a polymer of glucose. The glucose molecules in starches are joined together in the following manner:



It will be noted from this formula that the successive molecules of glucose are combined with each other by a *dehydration* process, which means that one glucose molecule loses a hydrogen ion and the next loses a hydroxyl ion. The hydrogen and hydroxyl ions combine to form water, and the two glucose molecules connect together at the points where the ions were removed.

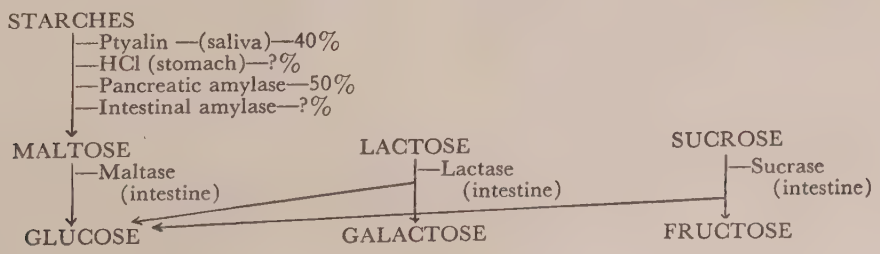
In addition to starches, another common source of carbohydrates is the disaccharides, which are combinations of only two molecules of monosaccharides. The common disaccharides in the diet are *maltose, sucrose,* and *lactose. Maltose* is a combination of two glucose molecules, and it is derived mainly by splitting starches into their disaccharide components, all of which are maltose. Sucrose is a combination of one molecule of glucose and one molecule of fructose. It is the same as cane sugar. Lactose is a combination of one molecule of glucose and one molecule of galactose. It is the sugar present in milk.

Digestion of Carbohydrates

Carbohydrate digestion breaks the starch or other carbohydrate polymers into their component monosaccharides. To do this, one molecule of water must be added to the compound at each point where two successive monosaccharides are joined. This is a process of *hydration,* which is opposite to the dehydration process by which the successive monosaccharides are combined with each other. The secretions of the digestive tract contain enzymes that catalyze this hydration process.

Scheme of digestion of carbohydrates. The schema below shows the digestion of the three most common carbohydrates in the diet, the starches, lactose, and sucrose. Starches and other large carbohydrates are digested principally by *ptyalin* in the saliva and *amylase* in the pancreatic juice, but perhaps to some extent by *hydrochloric acid* in the stomach and *intestinal amylase* in the small intestine. The resulting product of these reactions is the disaccharide maltose.

The intestinal secretions contain the enzymes *maltase, lactase,* and *sucrase,* which split maltose, lactose, and sucrose into their respective monosaccharides. The resulting products of carbohydrate digestion are glucose, galactose, and fructose as shown by the preceding schema. Because all the monosaccharides derived from maltose are glucose and half of those derived from the other two carbohydrates are glucose, it is evident that this substance is by far the most abundant end-product of carbohydrate digestion.



On the average, about 80 per cent of the monosaccharides formed by digestion is glucose, 10 per cent galactose, and 10 per cent fructose.

Absorption of Monosaccharides

Route of absorption. The monosaccharides are absorbed into the blood capillaries of many million small *villi* which are shown lining the intestinal wall in Figure 301A, and the structure of which is shown in Figure 301B. The capillaries of the villi empty into the portal venous system that flows into the liver, so that after absorption the monosaccharides must pass through the liver before entering the general circulation.

Mechanism of absorption. Monosaccharides are absorbed from the gastrointestinal tract by *active absorption.* The mechanism of this process is essentially the same as that for absorption from the tubules of the kidney, which was discussed in detail in Chapter 7. The monosaccharides combine with a carrier substance in the epithelial cells, and are transmitted in this combined form from the intestinal lumen to the opposite sides of the cells, where they are then released from the carrier into the capillary blood. For this process to occur, energy must be expended by the epithelial cells, which is the reason why the process is called "active" absorption. Active absorption is very important because it allows absorption to occur even when monosaccharides are present in the intestine in extremely small concentrations.

Fate of the Monosaccharides in the Tissues

Glucose in the blood and extracellular fluid—conversion of fructose and galactose to glucose. Immediately after the monosaccharides are absorbed from the gut, the glucose is rapidly transported throughout the body fluids in its present form. However, most of the fructose and galactose molecules are absorbed by the liver cells, converted into glucose, and then returned in this form to the blood, which means that these, too, are transported throughout the body in the form of glucose. Therefore, glucose is the basis of essentially all of the chemical reactions of carbohydrates in the body.

The concentration of glucose in the blood and extracellular fluid is approximately 90 mg. in each 100 ml., while the concentrations of fructose and galactose are usually very slight because of their rapid conversion to glucose.

Transport of glucose through the cell membrane—effect of insulin. Before glucose can be used by the cells it must be transported through the cell membrane. Unfortunately, the pores of the cell membrane are too small to allow glucose to enter by the process of diffusion. Here again it must be transported by an active process, the general principles of which are shown in Figure

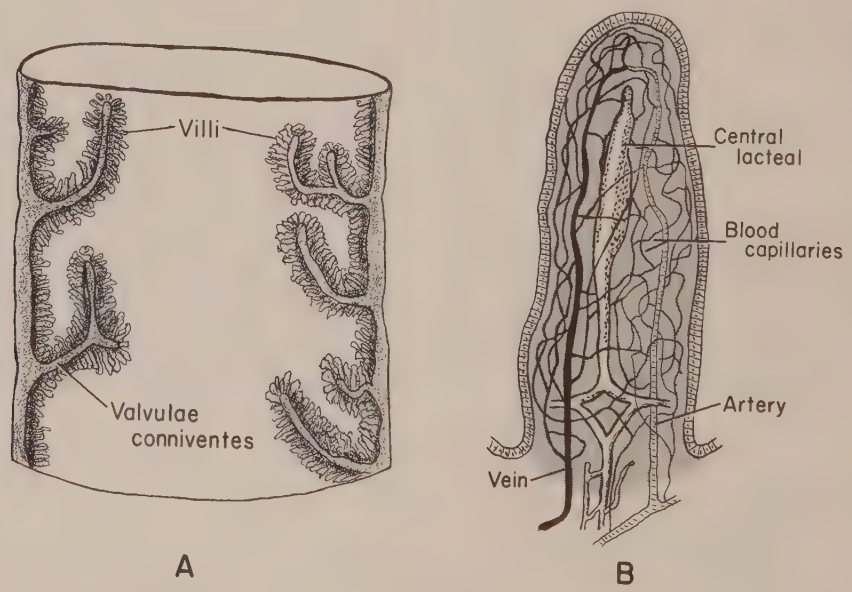


Figure 301. (A) Distribution of villi and mucosal folds on the inner surface of the small intestine. (B) Structure of a villus, showing the blood vessel system and the central lacteal.

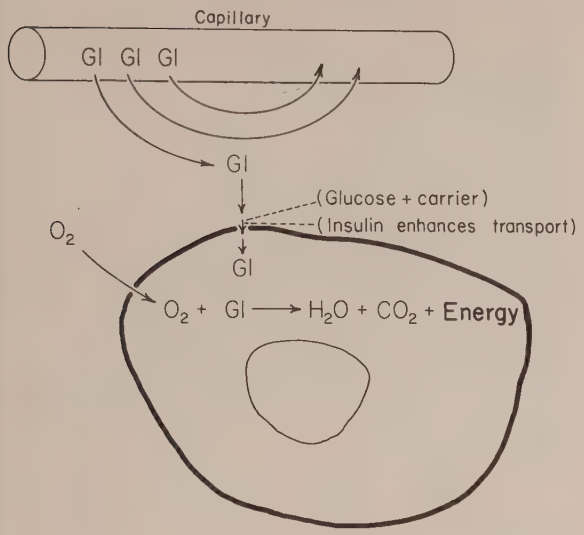


Figure 302. Transport of glucose from the capillary into the extracellular fluid and thence from the extracellular fluid into the cell to be utilized for energy.

302. Glucose first combines with a carrier, believed to be a phospholipid, in the cell membrane. Then it is transported to the inside of the cell where it breaks away from the phospholipid.

In some way that has not yet been completely explained, the hormone *insulin* greatly enhances the transport of glucose through the cell membrane. Some possible ways in which insulin might do this are: (1) by catalyzing the reaction between glucose and the carrier, (2) by removing glucose from the carrier on the inside of the cell, or (3) by forming one of the components of the carrier mechanism itself. Regardless of which of these might be correct, the rate at which glucose can be transported through the cell membrane is determined mainly by the amount of insulin available. When the pancreas fails to secrete insulin, 5 to 10 times less glucose enters than the cell needs. When insulin is secreted in excessive abundance, glucose enters so rapidly that glucose metabolism becomes much greater than normal. It is obvious, therefore, that the rate of carbohydrate metabolism is regulated in accordance with the rate of insulin secretion by the pancreas.

Regulation of blood glucose level. BUFFER EFFECT OF THE LIVER. After a meal, large quantities of monosaccharides are absorbed into the blood, and the glucose in the portal blood coming from the intestines rises from the normal level of 90 mg. per 100 ml. to as high as double this. However, this blood flows immediately into the liver, which removes about two thirds of the excess glucose before it can reach the general circulation. In this way the liver keeps the general blood concentration of glucose from ever rising above 120 to 130 mg. per 100 ml., despite very rapid absorption from the intestines.

The mechanism by which the liver removes the glucose from the portal blood is the following: Glucose is first absorbed through the cellular membranes into the liver cells under the influence of insulin. Then it is converted to *glycogen*, a polymer of glucose, and stored until a later time. When the blood glucose level falls to lower values several hours after a meal, the glycogen is split back into glucose, which is transferred out of the liver into the blood.

In essence, then, the liver is a buffer organ for blood glucose regulation, for it keeps the blood glucose level from rising too high and from falling too low.

EFFECT OF INSULIN PRODUCTION BY THE PANCREAS. After a person eats a large meal, the rise in blood glucose concentration stimulates the pancreas to produce large quantities of insulin. The insulin in turn promotes rapid transport of glucose into the cells, thus decreasing the blood glucose level back toward normal. Therefore, in addition to the liver buffer mechanism, this pancreatic production of increased quantities of insulin also aids in preventing excessive rises in blood glucose concentration.

EFFECT OF EPINEPHRINE AND SYMPATHETIC STIMULATION. A low blood glucose level stimulates the sympathetic centers of the brain, causing secretion of norepinephrine and epinephrine by the adrenal glands and excitation of all the sympathetic nerves throughout the body. Norepinephrine, epinephrine, and sympathetic stimulation all cause liver glycogen to split into glucose and then be emptied into the blood. This returns the blood glucose concentration back toward normal, acting as a protective mechanism against low blood glucose levels.

Gluconeogenesis. Another effect that occurs when the blood glucose level falls too low is the formation of glucose from proteins and to a much less extent from fats. This phenomenon is called *gluconeogenesis*. The importance of gluconeogenesis is that it provides glucose to the blood even during periods of starvation. Glucose, unfortunately, is not stored to a major extent in the body, for only 300 grams at most is stored in the form of glycogen in the liver and in skeletal muscle. Ordinarily this amount is not sufficient by itself to maintain the blood glucose concentration at normal values for more than 24 hours. However, as the blood glucose level falls below normal, gluconeogenesis begins and continues until an adequate supply of glucose is available again.

Later in this chapter it will be noted that most

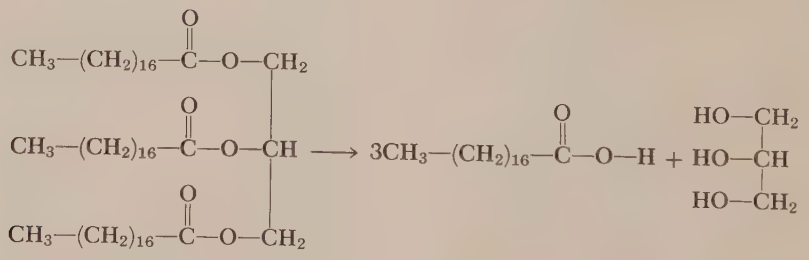
of the cells of the body can utilize fats for energy when glucose is not available. However, the neurons of the brain are unable to utilize fats, and without an adequate supply of glucose begin to die. This is the major reason why it is very important that the blood glucose concentration remain essentially normal even during long periods of starvation.

Energy from Glucose

Almost the only function of glucose in the body is to provide energy, though the glucose molecule is used rarely as a building stone for synthesis of other needed compounds. Energy is derived from glucose by two means: first, by splitting the molecule of glucose into smaller compounds, which liberates a small amount of energy, and, second, by splitting hydrogen ions away from the molecule and oxidizing these to form water, which liberates an extremely large amount of energy. These mechanisms of energy release are discussed in the following chapter.

DIGESTION, ABSORPTION, AND DISTRIBUTION OF NEUTRAL FATS

Neutral fats, like carbohydrates, are composed of carbon, hydrogen, and oxygen, though the relative abundance of oxygen in fats is considerably less than in carbohydrates. A representative molecule of neutral fat and its digestive end-products are the following:



It is evident from these formulas that a fat molecule contains two major components: first, a glycerol nucleus and, second, three fatty acid radicals. Each fatty acid radical is combined with the glycerol by a dehydration process, which was noted previously as the means by which monosaccharides also combine with each other to form complicated carbohydrates.

The differences among various fats lie in the composition of the fatty acids in the molecule. Most fats in the human body have fatty acids with 16 or 18 carbon atoms in their chains. The fats containing the longer chain fatty acids are a little more solid than those containing the shorter fatty acids. Other than this, the chemical and physical properties of most fats do not vary significantly from one to the other.

Some of the fatty acids in the body and in the diet are *unsaturated*, which means that at various points in the carbon chain the atoms are bonded together by double bonds rather than single bonds, and that there is a corresponding lack of two hydrogen atoms. The unsaturated fats are needed to form a few special structures of the cells, but otherwise even these perform the same functions as the saturated fats in providing energy for the metabolic processes.

Digestion of Fats

The digestion of fats, like that of carbohydrates, is a hydration process. It is catalyzed by enzymes called *lipases*, secreted in the stomach, pancreatic, and intestinal juices. The diagram below shows the complete schema of fat digestion.

$$\text{Fat} \xrightarrow{\text{(Gastric lipase + HCl)}} \begin{Bmatrix}\text{Fatty acids}\\ \text{Glycerol}\\ \text{Glycerides}\end{Bmatrix} \text{not over a few per cent}$$

$$\text{Fat} \xrightarrow{\text{(Bile + Agitation)}} \text{Emulsified fat}$$

$$\text{Emulsified Fat} \xrightarrow[\text{(Intestinal lipase)}]{\text{(Pancreatic lipase)}} \begin{Bmatrix}\text{Fatty acids}\\ \text{Glycerol}\\ \text{Glycerides}\end{Bmatrix} \begin{matrix}50\%(?)\\ \\ 50\%(?)\end{matrix}$$

Though a minute quantity of fat is digested in the stomach, most of it is digested by the pancreatic and intestinal lipases after it enters the small intestine. The *bile salts* from the liver aid greatly in this digestion by acting as a detergent which allows the mixing movements of the intestines to break the fatty globules of the food into small emulsified globules, thereby providing far more surface area on which the water soluble lipases can act. In the absence of bile salts, most of the fat passes on into the feces in an undigested state.

The end-products of fat digestion are *fatty acids, glycerol,* and *glycerides.* Glycerides are composed of a glycerol nucleus with one or two

of the fatty acid chains still attached. Though the end result of complete fat digestion is the splitting of the fats entirely into fatty acids and glycerol, the process goes to completion in only about 50 per cent of the fat molecules, leaving many glycerides still among the digestive products.

Because glycerides pass through the intestinal membrane with almost the same ease as glycerol and fatty acids, the digestive process is quite adequate for absorption to occur. In fact, a very small amount of finely emulsified neutral fat that has not been digested at all can even be absorbed through the membrane.

Absorption of the End-products of Fat Digestion

The end-products of fat digestion, like the monosaccharides, are also absorbed by the villi of the intestinal mucosa; but, unlike the monosaccharides, they are absorbed into the *central lacteal,* a lymphatic vessel, instead of into the blood. This is shown in Figure 301. As the glycerol, fatty acid, and glyceride molecules pass through the wall of the villus, they are reconverted into neutral fat. Thus, the digestion of fat into small chemical compounds allows it to *diffuse* through the intestinal wall, but, once through, the fat again becomes neutral fat as in the food.

Bile salts, in addition to their effect in helping to emulsify fat globules in the intestines, also aid the absorption of fats by performing a so-called *hydrotropic* action. This makes fatty acids and glycerides more soluble, allowing increased effectiveness of absorption.

There is a major difference between absorption of the digestive products of carbohydrates and those of fats, for the monosaccharides are absorbed almost entirely by active transport while the end-products of fat digestion are absorbed entirely by diffusion. The fatty substances become dissolved in the lipids of the epithelial membrane covering the villi and diffuse by this route through the cells, finally reaching the central lacteals of the villi.

Transport of fat through the lymphatics. Lymph is "milked" from the central lacteals into the abdominal lymphatics by rhythmic contraction of the villi. This contraction is caused by a hormone, *villikinin,* which is released from the intestinal mucosa when fats are in the chyme. After leaving the central lacteals, the neutral fat is transported upward through the *thoracic duct,* the major lymphatic channel of the body, to empty into the blood circulation at the juncture of the internal jugular and subclavian veins.

CHYLOMICRONS. The fat absorbed into the central lacteal becomes small fatty globules of about one micron in diameter called *chylomicrons.* These immediately adsorb proteins to their surface, which keeps them suspended in the lymph and keeps them from sticking to each other or to the walls of the lymphatics or blood vessels. It is in this form that fats are transported from the intestine, through the lymphatics, and into the blood.

After a fatty meal the level of chylomicrons in the circulating blood reaches a maximum in approximately two to four hours, sometimes becoming as much as 1 to 2 per cent of the blood, but within 2 to 3 hours almost all of them will have been deposited in the fat tissue of the body or in the liver.

Fat Tissue

Fat tissue is a special type of connective tissue that has been modified to allow the storage of neutral fat. It is found beneath the skin, between the muscles, between the various organs, and in almost all spaces not filled by other portions of the body. The cytoplasm of fat cells sometimes contains as much as 95 per cent neutral fat. These cells do not use the fat for any purpose other than to store it until it is needed elsewhere in the body.

Fat tissue provides a *buffer* function for fat in the circulating fluids. After a fatty meal the high concentration of fats in the blood is very soon lowered by deposition of the extra fat in the fat tissue. Then when the body needs fat for energy or other purposes it can be mobilized from the fat tissues and returned to the circulating blood. One of the means by which the body causes this release of fat is by secreting adrenocortical hormones, for these hormones increase the permeability of the fat cells and allow much of the fat to escape into the circulation. Because of this buffer function of the fat tissues, the fat in the fat cells is in a constant state of flux. Ordinarily half of it is removed every eight days, and new fat is deposited in its place.

Transport of Fats in the Body Fluids

Unesterified fatty acids. Most fat is transported in the blood from one part of the body to another in the form of *unesterified fatty acids,* which are a combination of free fatty acids with albumin, one of the plasma proteins. Every fat depository of the body is well supplied with an enzyme called *lipoprotein lipase* which is capable of splitting the neutral fat of the tissues into

glycerol and fatty acids. The fatty acids immediately combine with albumin in the blood and are then transported to other tissues of the body where they are released from the albumin. Some combine with glycerol in the fat depot areas to form new neutral fat. Others enter tissue cells where they are split into smaller molecules that are used to supply energy, as is explained below.

Lipoproteins. Lipoproteins are minute fatty particles covered by a layer of adsorbed protein. These are suspended in a colloid form in the plasma and to a less extent in other extracellular fluids. The *chylomicrons* are a type of lipoprotein because they are composed of lipid substances (neutral fat, phospholipids, and cholesterol) and a layer of adsorbed protein. However, in addition to the chylomicrons, large numbers of much smaller lipoprotein particles are also present in the blood. These are formed almost entirely in the liver, and their function is probably to transport neutral fat, phospholipids, and cholesterol from the liver to the different cells of the body.

Synthesis of Fat from Glucose and Proteins

Much of the fat in the body is not derived directly from the diet but instead is synthesized in the body. The fat cells themselves are capable of synthesizing a small amount of fat, but most of it is synthesized in the liver and then transported to the fat cells. Both glucose and amino acids derived from proteins can be converted into fat, but by far the most important source is glucose.

When an excess of glucose is in the diet and sufficient insulin is being secreted by the pancreas to cause all the glucose to enter the cells, almost all the extra glucose not used immediately for energy passes into the liver and is converted into fat. This is then transported by the blood to the fat tissues and deposited. Thus, the fat tissues provide a means for storing energy derived from carbohydrates as well as from fats. This conversion of other foods to fats explains why eating any type of food, whether it be fat, carbohydrate, or protein, can increase the amount of fat tissue.

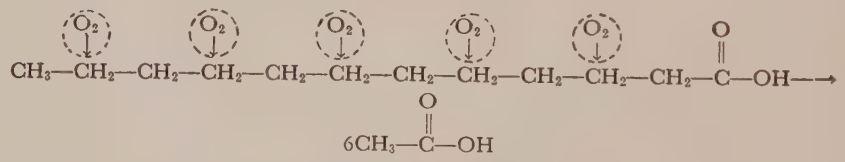
Functions of the Liver in the Utilization of Fat

The liver is undoubtedly the most important organ of the body for controlling fat utilization by the body. The liver, in addition to converting excess glucose into fat, converts fat into substances that can be used elsewhere in the body for special purposes. For example, before the tissue cells can make full use of the fats, some of the fats must be desaturated; some must be converted into the fatty substances, *cholesterol* and *phospholipids,* needed for cellular structures; and others must be broken into smaller forms before the cells can use them easily even for energy. The liver performs all these functions.

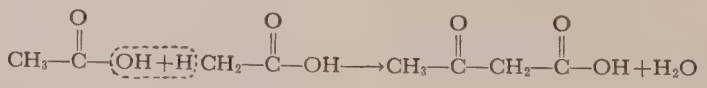
When the body is depending mainly on fats instead of glucose for energy, the quantity of neutral fat in the liver gradually increases. This is enhanced by adrenocortical hormones, for they cause the fat cells to mobilize their fat.

Energy from Fats

As the first stage in the utilization of fats for energy, the liver splits the neutral fat into glycerol and fatty acids. The glycerol, being very similar to some of the breakdown products of glucose, can then be used for energy in very much the same manner as glucose. However, by far the major amount of energy in the fat molecule is in the fatty acid chains, and before these can be used for energy they must be split into smaller chemical compounds. Ordinarily, this is accomplished by a chemical process called *alternate oxidation* of the carbon chain, which is illustrated by the following reaction:



The net result is the formation of many molecules of *acetic acid.* Then, immediately, pairs of acetic acid molecules *condense* to form *acetoacetic acid* in the following manner:



The acetoacetic acid is called a *keto acid,* and it in turn can change into several other closely related forms of keto acids, or even into acetone.

The keto acids are highly diffusible through cellular membranes. Therefore, as shown in Figure 303, the keto acids formed in the liver diffuse

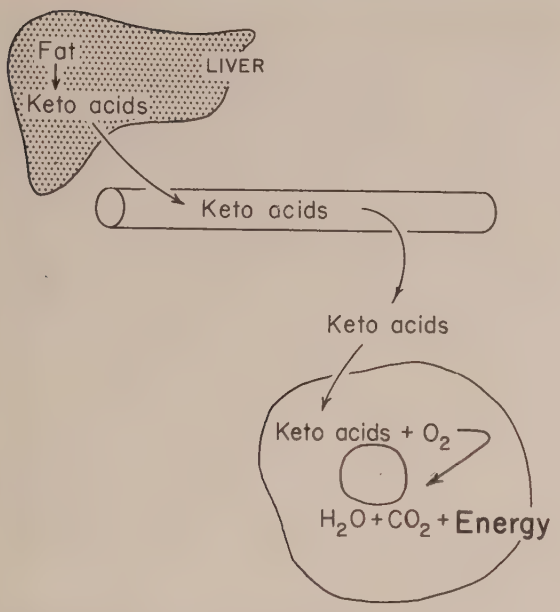


Figure 303. Formation and transport of keto acids from the liver to the cells, and utilization of keto acids for energy.

immediately into the blood and are transported to all cells throughout the body. There the keto acids are oxidized in the same manner as glucose to provide energy for cellular functions. This aspect of cellular metabolism is discussed in the following chapter.

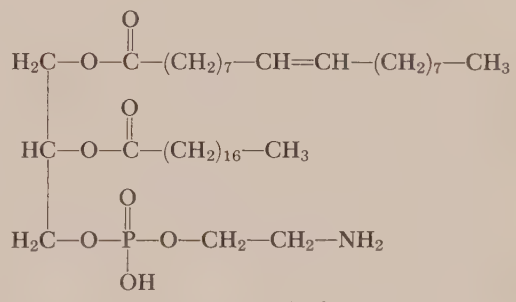
Small quantities of fatty acids can be split into acetic acid by the process of alternate oxidation in almost all cells of the body, and the resulting acetic acid molecules are utilized directly for energy. However, most of the fat is first split in the liver and then transported to the cells in the form of acetoacetic acid. This, in turn, is split into acetic acid that is used for energy.

Fat-sparing effect of carbohydrates. As long as sufficient glucose is available to supply the energy needs of the cells, it is burned in preference to the keto acids, and, if more than enough glucose is available, the excess is converted into fat. When glucose is available, the burning of fat stops; for this reason glucose is said to be a *fat sparer.*

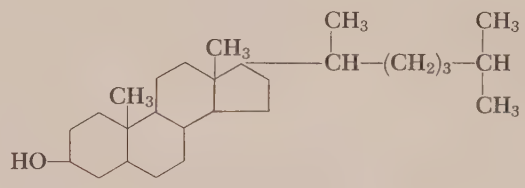
Effect of thyroid hormone on the amount of energy derived from fat. In addition to the effect of adrenocortical hormones in mobilizing fats, another hormone, thyroid hormone, greatly affects the utilization of fats by the body. Thyroid hormone increases the rates of all chemical reactions in the cells, and consequently increases the cellular demands for energy. Because all the glucose eaten each day is already either utilized directly by the cells or converted into fat, this extra energy must be derived from the stored fat, diminishing the amount of fat tissue in the body. For this reason thyroid hormone is frequently said to affect the metabolism of fat more than that of the other foods. Actually it increases the rate of metabolism of all foods, but because fat is the major means by which energy is stored in the body, it is the fat tissues that in the end are diminished.

Phospholipids and Cholesterol

Two additional substances, *phospholipids* and *cholesterol,* which have some physical properties similar to those of neutral fats, are present in large quantities throughout the body. The phospholipids are composed of glycerol, fatty acid, and a phosphate side chain. Cholesterol is composed mainly of a sterol nucleus that is synthesized from acetic acid as is also true of fatty acids. The chemical structures of these substances are the following:



A phospholipid



Cholesterol

Both phospholipids and cholesterol are fat soluble and are slightly water soluble. They are synthesized in all cells of the body, though to a much greater extent in the liver cells than in other cells. Both are transported in lipoproteins from the liver to other parts of the body.

Both phospholipids and cholesterol are major constituents of cell membranes and membranes of intracellular structures such as the nuclear membrane, the membranes of the endoplasmic reticulum, the membranes of mitochondria, of lysosomes, and so forth.

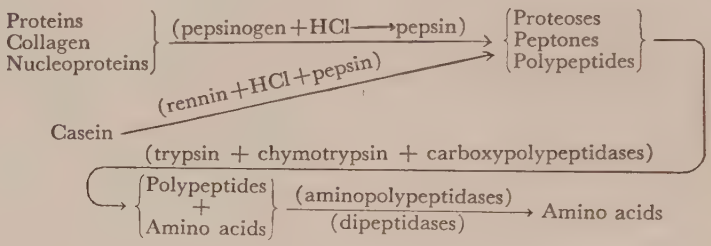
Some phospholipids are important in active transport of substances through membranes, acting as carriers with which the transported substances can combine and then diffuse through the membrane.

The precise function of cholesterol in the membranes is yet unknown, though the fact that it, like the phospholipids, is slightly water soluble as well as lipid soluble might make it a valuable adjunct to the transport of substances through cellular membranes.

Digestion of Proteins

Referring to the preceding equation, it is evident that a peptide linkage is another example of dehydration, which is the same means by which the component parts of fats and carbohydrates are combined. Therefore, the digestion of protein, like that of carbohydrates and fats, is accomplished by a process of hydration.

The following schema shows the sequence of protein digestion, which begins with *pepsin* action

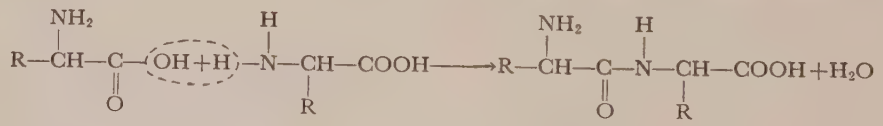


DIGESTION, ABSORPTION, AND DISTRIBUTION OF PROTEINS

Proteins are large molecules made up of many *amino acids* joined together. Amino acids, in turn, are small organic compounds that have an amino radical, $—NH_2$, and an acidic radical, —COOH, both on the same molecule. Twenty-three different amino acids are known to be present in the body proteins; the formulas of these are on page 343.

Thirteen of the amino acids can be synthesized in the body from other amino acids, but ten of them cannot. These ten are called *essential amino acids,* for they must be in the diet in order for the human body to form the proteins necessary for life.

Amino acids combine with each other to form proteins by means of *peptide linkages,* an example of which is illustrated by the following equation:



It will be noted that the product of the two combined acids, which is called a *peptide,* still has an amino radical and an acid radical, both of which can provide reactive points for combinations with additional amino acids. Most proteins contain several hundred to several thousand amino acids combined in this manner. The nature of the protein is determined by the types of amino acids in the protein and also by the pattern in which they are joined.

in the stomach. Pepsin is secreted into the stomach in the form of *pepsinogen,* a substance that has no digestive properties, but once it comes in contact with the hydrochloric acid of the stomach it is activated to form pepsin. The hydrochloric acid also provides an appropriate reactive medium for pepsin, for it can split proteins only in acid surroundings.

Protein is digested in the stomach into *proteoses, peptones,* and *polypeptides,* all of which are smaller combinations of amino acids than proteins—the proteoses are nearly as large as proteins, the peptones are intermediate in size, and the polypeptides are combinations of only a few amino acids. After entering the small intestine, these substances are further split by *trypsin* and *chymotrypsin* of the pancreatic juice into small polypeptides. Then the small polypeptides are finally split by *peptidases* of the pancreatic and intestinal juices into amino acids. Thus, the final products of protein digestion are the basic components of proteins, the amino acids.

Absorption of Amino Acids

Amino acids are absorbed from the gastrointestinal tract in almost exactly the same manner as monosaccharides, that is, by active processes. The precise nature of the active processes is not

known, but it has been shown that poisons that prevent metabolism in the intestinal mucosa block almost all absorption of amino acids. This is true also of absorption of monosaccharides.

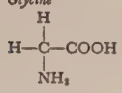
After absorption through the intestinal mucosa, the amino acids pass into the capillaries of the villi and thence into the portal blood, flowing through the liver before entering the general circulation.

Amino Acids in the Blood

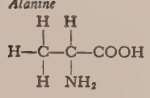
All of the different amino acids are in the blood and extracellular fluid in small quantities.

AMINO ACIDS

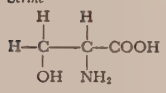
Glycine



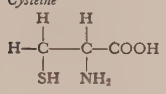
Alanine



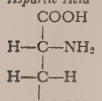
Serine



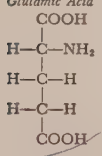
Cysteine



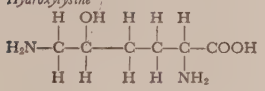
Aspartic Acid



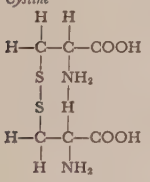
Glutamic Acid



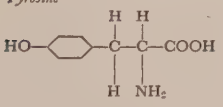
Hydroxylysine



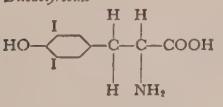
Cystine



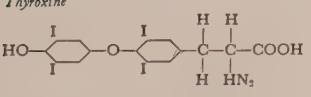
Tyrosine



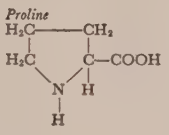
Diiodotyrosine



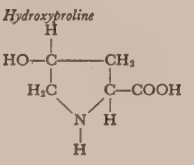
Thyroxine



Proline

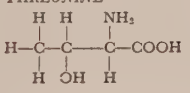


Hydroxyproline

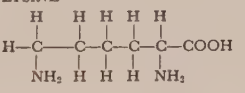


ESSENTIAL AMINO ACIDS

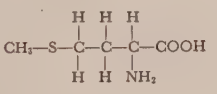
THREONINE



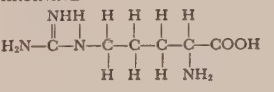
LYSINE



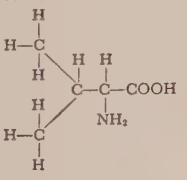
METHIONINE



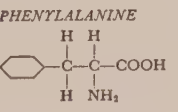
ARGININE



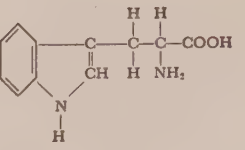
VALINE



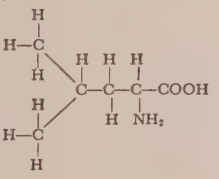
PHENYLALANINE



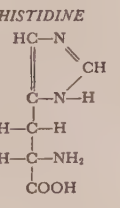
TRYPTOPHAN



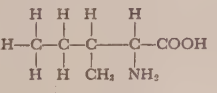
LEUCINE



HISTIDINE



ISOLEUCINE



However, their total concentration is only about 30 mg. in each 100 ml. of fluid; or, to express this another way, the total concentration of all the 23 different amino acids is only about one third that of glucose. The reason for this small concentration is that the amino acids, on coming in contact with cells, are absorbed very rapidly.

Buffer action of the liver and tissue cells for regulating blood amino acid concentration. The liver acts as a buffer for amino acids in the same manner that it acts as a buffer for glucose. When the concentration of amino acids rises very high, a large proportion of them is absorbed into the liver cells, where they can be stored in small quantities. When the amino acid concentration in the blood falls below normal, the stored amino acids pass back out of the liver cells into the blood to be used as needed elsewhere in the body.

Most other cells of the body also seem to have this ability to store amino acids to at least some extent and to release these into the extracellular fluid when its content of amino acids falls. As a result, amino acids are in a state of continual flux from one part of the body to another. If the amount of amino acids in the cells of one area falls too low, then amino acids will enter these cells from the blood and will be replaced by amino acids released from other cells. This continual flux of the amino acids among the various cells is illustrated by the diagram of Figure 304.

EFFECT OF ADRENOCORTICAL HORMONES ON AMINO ACID FLUX. Recent research has shown that adrenocortical hormones aid in the mobilization of amino acids from one area of the body to another. Though the precise nature of this mobilization function of adrenocortical hormones is yet unknown, it is believed that they increase the rate of active transport of amino acids through the cellular membranes and thereby promote their rapid transfer from tissue to tissue. When an area of the body is damaged and is in need of amino acids for repair of its cells, the rate of adrenocortical secretion becomes greatly increased, and the resultant mobilization of amino acids supplies the needed materials.

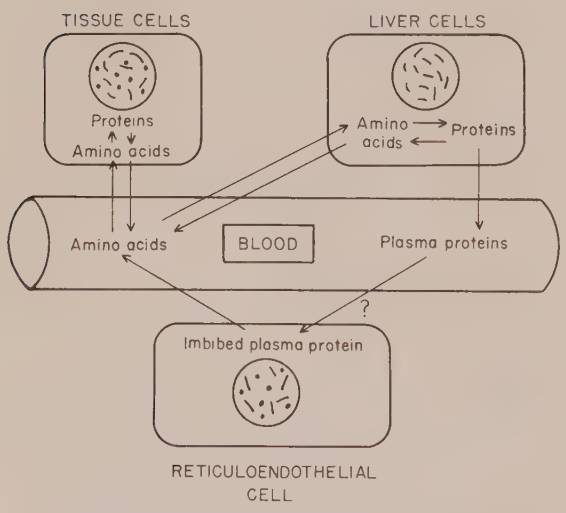


Figure 304. Flux of amino acids from one part of the body to another, and reversible equilibrium between the proteins of the different areas of the body.

The Tissue Proteins and Their Synthesis

The proteins of the cells perform two major functions. First, they provide most of the structural elements of the cells, and, second, they are the enzymes that control the different cellular chemical reactions. Therefore, the types of proteins in each cell determine its functions. Each cell is capable of synthesizing its own proteins, and this synthesis is controlled by the genes of the cell's nucleus in the manner described in Chapter 4. Basically, this process is the following:

Regulation of protein synthesis by the genes. The nucleus of each cell of the human body contains 46 chromosomes arranged in 23 pairs. Each of these chromosomes in turn is composed of many hundred *deoxyribose nucleoprotein* molecules connected together to form a long thread. It is believed that each of the molecules is a separate *gene*, and that its function is determined by its intrinsic chemical structure and also by its position in the chromosome thread.

Each gene of the nucleus controls the formation of a corresponding type of *ribose nucleoprotein* that is transported to the cytoplasm of the cell. This has a slightly different chemical composition from the deoxyribose nucleoprotein of the nuclear gene. All the respective ribose nucleoprotein molecules in turn act as enzymes to control the formation of the simple proteins in each cell. Since it is the simple proteins that perform the structural and enzymatic functions of the cell, the nuclear genes regulate in a roundabout way the entire function of the cell.

Formation of plasma proteins. The proteins in the plasma are of three different types: *albumin*, which provides the colloid osmotic pressure in the plasma; *globulins*, which provide the antibodies; and *fibrinogen*, which is used in the process of blood clotting. Almost all these are formed in the liver and then released into the blood, though a small portion, the globulins in particular, are formed by reticuloendothelial cells, plasma cells, and large lymphocytes. The plasma proteins

are the same as intracellular proteins except that they have been extruded into the circulating blood. The method of extrusion from the liver cells and plasma cells is unknown, but in the case of the large lymphocytes, it has been observed that the cellular membrane actually ruptures and empties its cytoplasmic contents into the lymph. Presumably some similar method is utilized by the liver and plasma cells.

Whenever the concentration of proteins in the blood falls too low to maintain normal colloid osmotic pressure, the production of plasma proteins by the liver increases markedly. Though the means by which this is effected is unknown, it obviously is of great value in maintaining normal circulatory dynamics, for, if the colloid osmotic pressure of the blood should vary either above or below normal, the transfer of fluids through the capillary membranes into and out of the tissue spaces would become abnormal.

Conversion of proteins into amino acids. All cells synthesize far more proteins than are absolutely necessary to maintain life of the cells. Therefore, if amino acids are needed elsewhere in the body, some of the cellular proteins can be reconverted into amino acids and then transported in this form. The reconversion process is catalyzed by enzymes, called *kathepsins,* that are in all cells. The quantity of proteins in a cell is determined by a balance between their rate of synthesis and their rate of destruction. Even the plasma proteins circulating in the blood are subject to reconversion into amino acids, for they can be imbibed by reticuloendothelial cells and perhaps other cells and then split into amino acids by the intracellular enzymes.

The constant balance between amino acids and proteins in the cells, and between amino acids and the plasma proteins, is shown in Figure 304. By this constant interchange of amino acids, the proteins in all parts of the body are maintained in reasonable equilibrium with each other. If one tissue suffers loss of proteins or if the blood suffers loss of plasma proteins, many of the proteins in the remainder of the body will soon be converted into amino acids, which are transported to the appropriate point to form new protein. For example, in widespread cancer that is using extreme quantities of amino acids for formation of new cancer cells, the amino acids are derived continually from the tissue proteins, leading to serious debility. Also, when large quantities of blood are lost, the plasma proteins are replenished back to normal within approximately 5 to 7 days by the transfer of amino acids from the tissue proteins to the liver where new plasma proteins are formed.

Use of Amino Acids to Synthesize Needed Chemical Substances

Most metabolic reactions in the cells require special chemical substances to keep them operating. For instance, the muscles require large quantities of adenine and creatine to cause muscular contraction, the blood cells require large amounts of heme to form hemoglobin, and the kidneys require large quantities of glutamine to be used in forming ammonia; these are but a few of the special substances needed by various tissues to perform their functions. Most of these substances are synthesized from amino acids.

Also, many of the hormones secreted by the endocrine glands are synthesized from amino acids. These include epinephrine and thyroxine which are synthesized from tyrosine, histamine synthesized from histidine, several pituitary hormones that are themselves large proteins, parathyroid hormone which is a small protein, and insulin which is a very large polypeptide.

Derivation of Energy from Amino Acids

In addition to the use of amino acids for synthesizing new proteins or other chemical substances, some of them are also used for energy, as shown in Figure 305. The first step in using proteins for energy is to remove the amino radical. This process is called *deamination,* and it *occurs in the liver.* The removed amino radical is converted into ammonia, which in turn combines with carbon dioxide to form urea. This is excreted by the kidneys into the urine. Once again the liver is extremely important for one of the metabolic processes.

Referring back to the formulas of the amino acids, it will be evident that removal of the amino radical from certain of these acids still leaves

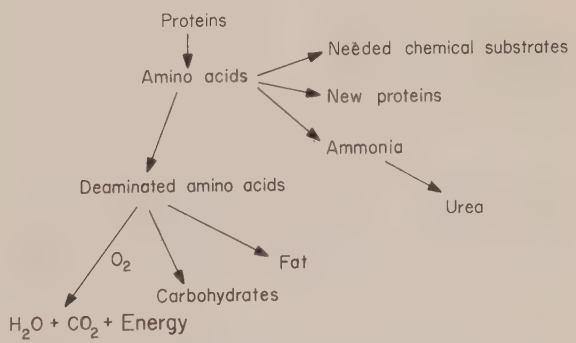


Figure 305. Complete schema for utilization of proteins in the body.

very complicated chemical compounds. Some of these cannot then be utilized by the body because of their nature, and, therefore, are excreted directly into the gastrointestinal tract and lost in the feces. At least 60 per cent of the deaminated amino acids, however, have sufficiently simple chemical structures so that they can enter into the same cellular reactions as glucose and keto acids. These are often directly oxidized to form water and carbon dioxide, liberating energy in the process, or, if energy is not needed at the moment, they can be converted into fat or carbohydrate and later utilized for energy in the form of keto acids or glucose.

Conversion of proteins to fats or carbohydrates. Ordinarily the body's cells must synthesize about 45 gm. of new proteins each day to replace the proteins being destroyed by the natural processes of wear and tear. If extra quantities of amino acids above the amount needed for this purpose are eaten, these normally are deaminated and converted to fats or carbohydrates, or are used for energy. In other words, even normally, excess proteins are not stored in the body in the form of proteins.

Protein-sparing effect of carbohydrates and fats during starvation. When an insufficient quantity of protein is eaten, the major portion of the energy needed for the chemical processes of the cells is derived from carbohydrates and fats as long as these are available. This is called the *protein-sparing* effect of these substances. However, when the stores of these have been greatly depleted, amino acids begin to be mobilized and deaminated to be used for energy. One can live for another few days on this energy derived from the proteins, but this process rapidly depletes the cells of their functional elements and soon leads to death.

ABSORPTION OF ELECTROLYTES AND WATER

Absorption of electrolytes. Electrolytes are absorbed from the gastrointestinal tract in almost exactly the same manner as from the tubules in the kidneys as described in Chapter 7. Sodium, for instance, is *actively absorbed;* that is, it combines with a carrier in the epithelial cells and is transported through the intestinal membrane in this form to be released on the opposite side into the blood.

Though less definitive experiments are available for absorption of other electrolytes from the gastrointestinal tract, it is believed that potassium, calcium, magnesium, chloride, phosphates, and iron are all also actively absorbed in a similar manner.

Absorption of water. Water, like fats, is absorbed by *diffusion.* This means simply that random motion of the water molecules eventually carries them through the epithelial pores into the extracellular fluid.

Water absorption is controlled almost entirely by crystalloidal osmotic forces that operate as follows: When the monosaccharides, amino acids, and electrolytes are absorbed from the small intestine by active absorption, the crystalloidal osmotic pressure of the intestinal fluids becomes very slight because of the loss of the crystalloids. On the other hand, the crystalloidal osmotic pressure of the interstitial fluid on the opposite side of the epithelial membrane becomes increased. As a result an osmotic pressure gradient develops which causes water to be absorbed by osmosis from the intestinal lumen into the extracellular fluids. It is in this way that eight or more liters of gastrointestinal fluid are absorbed from the gastrointestinal tract each day.

REFERENCES

Brauer, R. W.: Liver circulation and function. *Physiol. Rev.* 43:115, 1963.

Crane, R. K.: Intestinal absorption of sugars. *Physiol. Rev.* 40:789, 1960.

Dole, V. P., and Hamlin, J. T., III: Particulate fat in lymph and blood. *Physiol. Rev.* 42:674, 1962.

Fritz, I. B.: Factors influencing the rates of long-chain fatty acid oxidation and synthesis in mammalian systems. *Physiol. Rev.* 41:52, 1961.

Hogben, C. A. M.: The alimentary tract. *Ann. Rev. Physiol.* 22:381, 1960.

Meister, A.: *Biochemistry of the Amino Acids.* New York, Academic Press, 1957.

Olson, R. E., and Vester, J. W.: Nutrition-endocrine interrelationships in the control of fat transport in man. *Physiol. Rev.* 40:677, 1960.

Pigman, W.: *The Carbohydrates.* New York, Academic Press, 1957.

Randle, P. J.: Endocrine control of metabolism. *Ann. Rev. Physiol.* 25:291, 1963.

Springall, H. D.: *The Structural Chemistry of Proteins.* New York, Academic Press, 1954.

White, A., Handler, P., Smith, E., and Stetten, D., Jr.: *Principles of Biochemistry.* 2nd Ed. New York, McGraw-Hill Book Company, Inc., 1959.

RELEASE OF ENERGY FROM FOODS, AND NUTRITION

31

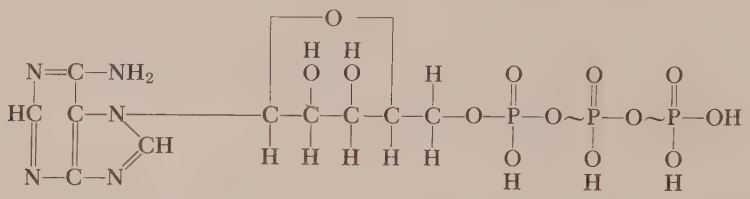
The major function of all the digestive and metabolic processes of the body is to provide energy for performing the various bodily functions. Energy is required to lift an arm, to move a leg, or to do any activity employing muscular contraction. It is needed for the secretion of digestive juices, the development of membrane potentials in nerves and other cells, the synthesis of new chemical compounds, and active absorption of substances from the gastrointestinal tract or kidney tubules. In short, almost all functions performed by the body require energy that in turn must be supplied by the ingested food. The digestion, absorption, and intermediary steps in preparing the food for energy release were discussed in the previous chapter. The final steps for release of energy from the foods—that is, the end stages of metabolism—are described in the present chapter.

ADENOSINE TRIPHOSPHATE AS THE COMMON PATHWAY OF ALL ENERGY

The cells do not use the actual foods for their immediate supply of energy. Instead, they use a chemical compound called *adenosine triphosphate* (ATP) for this energy, and the foods are

then used to synthesize more adenosine triphosphate after it has been used. The importance of this compound to the function of the cell was pointed out in Chapter 3. In the present chapter is explained the role of adenosine triphosphate in the overall utilization of energy by the body.

The formula for adenosine triphosphate is the following:

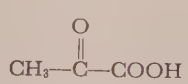


Extremely large amounts of energy are stored in this molecule at the *bonds* where the last two phosphate radicals join with the remainder of the molecule. These bonds (~) are called *high energy phosphate bonds.* Every time a cell needs energy a phosphate radical is broken away from adenosine triphosphate at the high energy bond, and this liberates the needed energy. Each mole of adenosine triphosphate releases 7000 calories of energy for each of the two high energy bonds that is broken. In short, there is a storehouse of adenosine triphosphate in each cell that provides the necessary energy for muscular contraction, development of membrane potentials, active absorption, active secretion, synthesis of chemical compounds, and other functions performed by the cells, but this adenosine triphosphate must be replenished continually.

Formation of Adenosine Triphosphate

Use of energy from carbohydrates to form adenosine triphosphate. In the preceding chapter it was noted that carbohydrates are digested to form glucose, or are changed into glucose after absorption. Then the glucose is used by the cells for energy. Part of the energy is released from glucose by a process called *glycolysis* that does not require oxygen, but by far the major amount of energy is released when the glucose is oxidized.

GLYCOLYSIS. In glycolysis, the glucose molecule, which has six carbon atoms, is split by a series of cellular enzymes into two smaller molecules having only three carbon atoms. Then the three-carbon-chain molecules are modified to form *pyruvic acid,* which has the following formula:



During glycolysis a small amount of energy is released from the glucose molecule, and this energy is used to form adenosine triphosphate. By splitting glucose to form two pyruvic acid molecules, energy is liberated without the expenditure of any oxygen. This is illustrated by the first stage of the reaction in Figure 306.

OXIDATIVE RELEASE OF ENERGY FROM CARBOHYDRATES. After the glucose has been split into pyruvic acid molecules, these can then be metabolized with oxygen to form carbon dioxide and water. This reaction is shown by the second stage in Figure 306. The oxidative metabolism of pyruvic acid provides about eighteen times as much energy as the glycolytic breakdown of glucose to form pyruvic acid. Therefore, by far the major amount of energy liberated from carbohydrates for the performance of cellular function is derived from *oxidative metabolism.*

The precise chemical reactions by which pyruvic acid is split into smaller molecules and oxidized have been worked out in great detail, the general principles of which are shown in Figure 307. The reactions of the first stage, called the *tricarboxylic acid cycle,* split the pyruvic acid molecule into carbon dioxide and hydrogen; the carbon dioxide is removed by enzymes called *decarboxylases,* and hydrogen atoms are removed by *dehydrogenases.* In the second stage, called *oxidation,* the hydrogen reacts with oxygen to form water.

When hydrogen atoms are split away from pyruvic acid by the dehydrogenases, they immediately combine with a substance called *coenzyme.* Then, under the influence of other enzymes, hydrogen atoms are passed to *flavoprotein mole-*

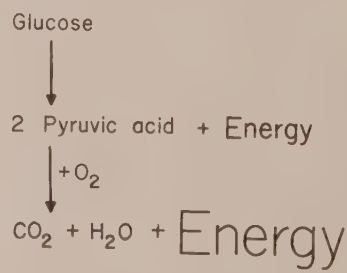


Figure 306. Derivation of energy from glucose by glycolysis and by oxidation.

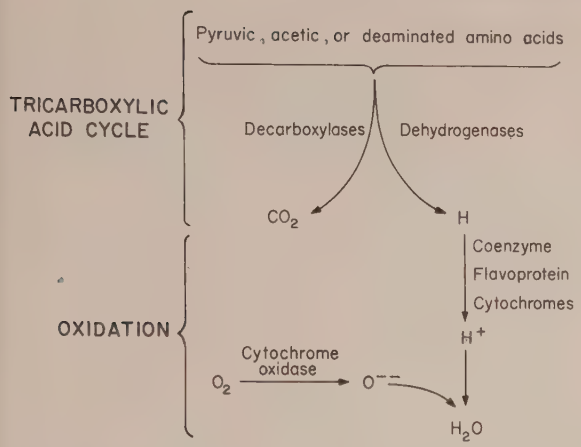


Figure 307. Splitting of pyruvic acid, acetic acid, or deaminated amino acids into carbon dioxide and hydrogen in the tricarboxylic acid cycle, and oxidation of the released hydrogen atoms by the cellular oxidative enzymes.

cules and finally to *cytochrome molecules*. At this point the hydrogen atoms are released into the surrounding fluid as hydrogen ions. Simultaneously, dissolved oxygen that has been carried to the tissues by hemoglobin is changed into oxygen ions by *cytochrome oxidase*. Obviously, the presence of ionic hydrogen and ionic oxygen in the same solution provides two highly reactive substances that immediately form water molecules. Thus, the hydrogen atoms removed from the pyruvic acid become oxidized with oxygen to form water. During the various stages for oxidizing hydrogen, energy is released to form still more adenosine triphosphate molecules.

One might ask why it is necessary for the hydrogen and oxygen to go through the complicated stages of the above reactions, for it is well known that hydrogen and oxygen can combine with each other very rapidly simply by being burned together. The answer to this question is that the indirect procedure is required to channel the released energy in the proper direction to form new adenosine triphosphate.

The quantity of adenosine triphosphate formed in the different stages of glucose metabolism is the following: For each molecule of glucose metabolized, 2 molecules of adenosine triphosphate are formed during glycolysis, 2 are formed in the tricarboxylic cycle, and 34 are formed during oxidation of hydrogen, making a total of 38 molecules of adenosine triphosphate for each molecule of glucose metabolized.

The total amount of energy in each mole of glucose is 686,000 calories. Of this amount, 266,000 become stored in the form of ATP. The remainder is lost as heat caused by the chemical reactions. Thus, the overall *efficiency* of energy transfer between glucose and ATP is 39 per cent, the remaining 61 per cent of the energy becoming heat, which represents waste.

Use of energy from fats and proteins to form adenosine triphosphate. It was pointed out in the preceding chapter that fatty acids are split into keto acids, and that amino acids derived from proteins are deaminated to form deaminated amino acids. When the keto acids enter tissue cells they split further into acetic acid molecules, and the deaminated amino acids that can be used for energy are also changed into relatively small two- or three-carbon-chain molecules. Then the same decarboxylases and dehydrogenases that remove carbon dioxide and hydrogen from pyruvic acid do the same for the acetic or deaminated amino acids, and the hydrogen atoms are oxidized as explained above for the carbohydrates. Large amounts of energy are released, especially during the oxidation of the hydrogen atoms, to synthesize adenosine triphosphate.

Regulation of cellular oxidation by adenosine diphosphate. The oxidation reactions that result in the formation of adenosine triphosphate cannot proceed unless adenosine diphosphate is available from which the adenosine triphosphate can be formed. Therefore, the oxidative breakdown of foodstuffs is controlled by the presence or absence of adenosine diphosphate. Every time adenosine triphosphate is used by the cells for energy it becomes adenosine diphosphate, and the newly formed adenosine diphosphate immediately initiates reactions to cause release of new energy from the foodstuffs. Then, when all the adenosine diphosphate has been resynthesized into adenosine triphosphate, the synthetic processes cease.

Interaction of Adenosine Triphosphate with Phosphocreatine

Another substance that contains high energy phosphate bonds, phosphocreatine, is also present in the cells, in quantities several times as great as those of adenosine triphosphate. When adenosine triphosphate is available in large amounts, much of it is used to form phosphocreatine in accordance with the reactions shown in Figure 308, and as rapidly as it is used more adenosine triphosphate is formed. This results in the build-up of large quantities of both adenosine triphosphate and phosphocreatine. When the cell demands energy rapidly and in large amounts, energy is released from adenosine triphosphate directly to the functional elements of the cells. Then the energy stored in the phosphocreatine is used immediately to reform new adenosine triphosphate. This

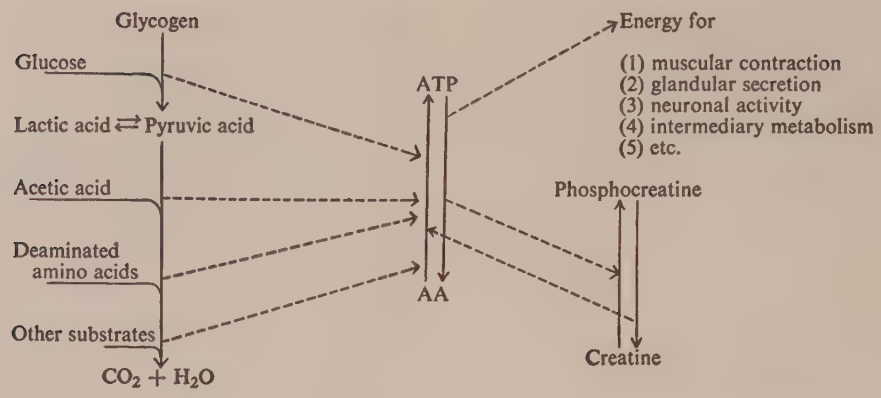


Figure 308. Overall schema for transfer of energy from the food to the functional elements of the cells.

reaction occurs much more rapidly that the oxidative reactions, and provides a very rapid source of extra energy that can be used to keep the cell functioning at a high rate of metabolism for a short period of time, even though the oxidative methods for providing energy are slow to respond.

An overall schema for the energy mechanisms of the cell. Figure 308 illustrates an overall schema for the chemical reactions that provide energy for cellular functions. This schema shows that the breakdown of adenosine triphosphate to adenosine diphosphate and adenosine monophosphate (AA in schema) releases the needed energy for muscular contraction, glandular secretion, neuronal activity, intermediary metabolism, and other energy functions of the cell. Within a few seconds, phosphocreatine also breaks down, providing energy for resynthesis of much of the adenosine triphosphate. During the ensuing minutes, the adenosine triphosphate and phosphocreatine are resynthesized by energy removed from the foodstuffs, part of it by glycolysis and other anaerobic procedures, but 90 per cent or more by the oxidation of pyruvic acid, acetic acid, deaminated amino acids, and a few other substances such as alcohol, glycerol, and lactic acid.

The oxygen debt. For a few seconds at a time one can perform very strenuous feats of exercise requiring energy release many times that which can be sustained over a long period of time. This immediate burst of energy is provided mainly by the adenosine triphosphate and phosphocreatine stored in the cells. During the next few minutes, while the adenosine triphosphate and phosphocreatine are being resynthesized by the oxidative processes of metabolism, an extra quantity of oxygen must be utilized. Also, some of the stores of oxygen in the hemoglobin of the blood, in the myoglobin of muscle, and in the dissolved state in the body fluids will have been used during the rapid burst of energy, and this, too, must be replenished after the exercise is over. The extra oxygen that must be used to restore completely normal conditions after exercise is over is called the *oxygen debt.*

In essence, the ability to develop an oxygen debt explains why an athlete continues to breathe very hard for many minutes after running a race.

NUTRITION

The term *nutrition* means the supplying of foods that are required to keep one alive and healthy. These foods include carbohydrates and fats, which supply most of the body's energy, and proteins, vitamins, and minerals, which are required for synthesis of special structures and special chemical compounds needed by the body.

Foods Supplying Energy

Peoples in various parts of the world have widely differing diets, and some even obtain different proportions of their energy from the various types of foodstuffs in comparison with others. Approximately 45 per cent of the energy of the average American is derived from carbohydrates, 40 per cent from fats, and 15 per cent from proteins. In other less prosperous parts of the world, the abundance of fats and proteins in the diets is often less than half these values, while the energy derived from carbohydrates sometimes rises to as much as 80 per cent of the total.

The calorie as a measure of energy. The energy of a food is measured in terms of the amount of heat that is liberated by the complete breakdown of the food into end-products, and this is

expressed in *Calories,* a unit for measuring heat. One Calorie is the amount of heat required to raise the temperature of 1 kilogram of water 1 degree Centigrade.

The Calorie is a very good measure of food energy, for most of the energy released in the body eventually becomes heat anyway. For example, the chemical reactions for extracting energy from the foods are so inefficient that 75 to 90 per cent of the released energy becomes heat even before any cellular functions are performed because the chemical reactions themselves evolve this much heat. The remaining 10 to 25 per cent is converted into muscular action or other functional activities of the body, but after these functions are performed most of the energy still becomes heat. As an example, a large amount of energy is used to pump blood around the circulatory system. Then, as the blood flows through the circulatory system, the energy that has been imparted to the blood is converted into heat because of friction between the blood and the walls of the vessels. In this way, all the energy expended by the heart eventually becomes heat. Likewise, almost all the energy expended by the skeletal muscles of the body eventually becomes heat, because most of it is used to overcome the friction of the joints and the viscosity of the tissues, and these two effects in turn convert the energy into heat.

Energy content of the different types of foods. The amount of energy released to the body by 1 gm. of each of the three types of foods respectively is the following:

	Calories
Carbohydrate	4.1
Fat	9.3
Protein	4.1

From the figures shown here it is evident that 1 gm. of fat supplies more than twice as much energy as 1 gm. of either carbohydrate or protein. For this reason, fats in the diet are very deceptive. Often one thinks that he is eating a relatively non-fat diet, and yet may be obtaining as much energy from fat as from carbohydrates. A second reason why fat is deceptive is that it often occurs in a pure form in foods, while carbohydrates and proteins are generally diluted several-fold with water. When one eats potatoes with butter, the fat of the butter generally contains almost as much energy as the entire helping of potato, for two reasons: First, the butter itself supplies 2¼ times as much energy per gram as the starches, and, second, the starches comprise only about one-sixth the bulk of the potato, because most of it is water.

Protein Requirements of the Body

However low the concentration of amino acids in the blood, the liver continually deaminates some of them all the time, and these either are excreted or are used for energy. Because of this continual loss of amino acids, the proteins of the body diminish constantly unless they are replenished by proteins in the diet. Normally, approximately 45 gm. of protein are required each day to replace this continual loss.

Partial proteins. Certain types of proteins do not contain all the essential amino acids in the proportion in which they are needed in the body. When one eats such a protein his requirement will be considerably above 45 gm. per day. For instance, the proteins from vegetables or grains are composed of ratios of the various amino acids different from those found in the human body. On the other hand, the proteins of animal origin, in general, have almost exactly the same amino acid compositions as those of the human being. Therefore, the person who requires 45 gm. of animal protein in his diet each day might require as much as 65 gm. of vegetable protein, the exact amount depending on the type of vegetable eaten. The proteins that cannot supply the right proportions of the different amino acids are called *partial proteins,* because they supply only part of all the needed amino acids.

Special Need for Unsaturated Fat in the Diet

A small amount of highly saturated fat is essential for nutrition of animals, and the body cannot desaturate fat sufficiently by its own metabolic processes to supply it. The types of unsaturated fatty acids usually available in the diet are *arachidonic, linoleic,* and *linolenic* acids. These are believed to be needed by the cells to form cellular structures, for without them animals develop skin sores, mental changes, and other evidences of general cellular debility. Whether or not these same effects would occur in the human being is not known, because such a small amount of these substances is required in the diet that no human being has ever been known to suffer from a deficiency of them.

Daily Energy Requirements

An average man of 70 kilograms who lies in bed all day long and does nothing else except eat and exist usually requires about 1850 Calories of energy each day. If he simply sits in a chair, another

200 or more Calories are required. Therefore, about 2000 Calories per day is the normal basal amount of energy required simply for living. In addition, any type of exercise requires still more energy, which is shown by the values in Table 7 for different types of activity. From this table it is evident that walking upstairs requires approximately 17 times as much energy as lying in bed asleep. However, this tremendous rate of energy utilization can be continued only for short periods of time. Over long periods, a well conditioned worker can average as much as 6000 to 8000 Calories of energy expenditure each 24 hours, or, in other words, as much as four times the basal rate.

Vitamin Requirements of the Body

The vitamins are chemical compounds needed in only minute quantities by the body to perform special functions. The daily requirements of each of the vitamins is given in Table 8, and the vitamin content of different foods is given in Table 9. In general, eating a balanced diet will provide an adequate quantity of all the different vitamins. Occasionally, though, an abnormality makes it impossible to utilize one of the vitamins, in which case a vitamin deficiency disease can occur even in the presence of a normally satisfactory diet.

Table 7. *Energy Expenditure per Hour during Various Types of Activity for a 70 Kg. Man (M. S. Rose)*

Form of Activity	*Calories per Hour*
Sleeping	65
Awake lying still	77
Sitting at rest	100
Standing relaxed	105
Dressing and undressing	118
Tailoring	135
Typewriting rapidly	140
"Light exercise"	170
Walking slowly (2.6 miles per hour)	200
Carpentry, metal working, industrial painting	240
"Active exercise"	290
"Severe exercise"	450
Sawing wood	480
Swimming	500
Running (5.3 miles per hour)	570
"Very severe exercise"	600
Walking very fast (5.3 miles per hour)	650
Walking up stairs	1100

Table 8. *Daily Requirements of the Vitamins*

Vitamin	*Daily Requirement*
A	5000 units
Thiamin	1.8 mgm.
Riboflavin	1.8 mgm.
Niacin	18.0 mgm.
Ascorbic acid	75.0 mgm.
D (children and during pregnancy)	400 units
E	unknown
K	none
P	unknown
Folic acid	0.5 mgm. (?)
B_{12}	2.0 μgm. (?)
Inositol	unknown
Pyridoxine	2.0 mgm. (?)
Pantothenic acid	unknown
Biotin	unknown
Para-aminobenzoic acid	unknown

The precise physiologic functions of most of the vitamins in the body are not known, but from the physical effects caused by lack of them in the diet, their functions can at least be speculated upon as follows.

Vitamin A. Vitamin A is used by the eyes to synthesize the light-sensitive retinal pigments used by the rods and cones for vision. This was discussed in detail in Chapter 27. Also, lack of vitamin A causes the epithelial structures, such as the skin, the intestinal mucosa, and the germinal epithelium of the ovaries and testes, to become highly *keratinized,* or horny. This leads to scaliness of the skin, failure of growth of young animals, failure of reproduction, and even hardening of the cornea, causing corneal opacity and blindness.

Thiamine. Thiamine forms a compound in the cells called *thiamine pyrophosphate,* which is probably part of a decarboxylase that removes carbon dioxide from pyruvic acid and other substances. Without thiamine the oxidative processes for releasing energy from foodstuffs become deficient, and this can cause almost any type of abnormality in the body. It especially affects the functions of the nervous system, the heart, and the gastrointestinal system.

Lack of thiamine causes pathologic changes in neurons and in the myelin sheaths of nerve fibers, often resulting in actual destruction of the cells or irritation or degeneration of peripheral nerves. Frequently the fibers of peripheral nerves become so irritable that a condition called *polyneuritis* results, which manifests itself by excruciating pain along the course of the nerves.

Table 9. *Vitamin Content of Various Foods (Modified from W. H. Eddy and G. Dalldorf).*

Vitamin A units/100 gm.	*Thiamine mgm./100 gm.*	*Riboflavin mgm./100 gm.*	*Niacin mgm./100 gm.*	*Vitamin C mgm./100 gm.*
Apricots 5000	Barley 0.450	Dried beans 0.30	Asparagus 1.2	Asparagus 45
Broccoli 4000	Dried beans 0.540	Almonds 0.50	Barley 4.7	Brussel sprouts 95
Butter 3500	Buckwheat 0.450	Beef 0.20	Soybeans 4.0	Cabbage 65
Carrots 6000	Dried cowpeas 0.900	Cheese 0.55	Dried beans 3.0	Butter 70
Collards 8000	Egg yolk 0.320	Chicken 0.30	Beef 5.4	Cauliflower 82
Chard 9000	Whole wheat 0.585	Collards 0.25	Chicken 8.5	Lemon juice 53
Endive 15,000	Pork meat 1.0	Eggs 0.40	Collards 2.3	Mustard greens 115
Kale 20,000	Lamb meat 0.330	Lamb 0.24	Corn 1.3	Orange juice 45
Liver 70,000	Beef meat 0.120	Liver 2.60	Lamb 8.0	Fish 20
Mustard greens 10,000	Liver 0.400	Milk 0.18	Liver 17.0	Potatoes 10
Sweet potatoes 3000	Millet 0.700	Peanuts 0.45	Mackerel 2.1	Lean meats 5
Pumpkin 7000	Brown rice 0.370	Pork 0.24	Pork 7.0	Liver 25
Spinach 25,000	Wheat germ 2.0	Grains 0.10	Salmon 6.2	
			Wheat 4.3	

	Pantothenic Acid mgm./100 gm.	*Pyridoxine mgm./100 gm.*	*Inositol mgm./100 gm.*	*Biotin mgm./100 gm.*	*Folic Acid mgm./100 gm.*
Meats	0.8	0.1	60	0.007	0.15
Eggs	1.4	0.02	33	0.009	0.90
Milk	0.3	0.13	18	0.005	0.005
Cereals	0.4	0.05	65	0.005	0.10
Fruits	0.2	0.05	95	0.003	0.05
Vegetables	0.5	0.15	70	0.007	0.08

In the heart, thiamine deficiency decreases the strength of the muscle. The heart becomes greatly dilated and pumps with little force, resulting in *congestive heart failure.*

In the gastrointestinal tract, thiamine deficiency causes weakness of the intestinal muscle, poor secretion of digestive juices, and poor maintenance of the intestinal mucosa. Severe indigestion, constipation, lack of appetite, and other symptoms very often develop.

A thiamine deficient person who has neuritis, enlargement of the heart, and gastrointestinal symptoms all at the same time is said to have *beriberi.*

Niacin and riboflavin. Niacin and riboflavin are utilized in the body to form respectively coenzyme and flavoprotein. Earlier in the chapter it was pointed out that these two substances are required during the oxidation of hydrogen to combine with hydrogen atoms after they are removed from pyruvic acid and other substrates. Without niacin and riboflavin, therefore, the oxidation of foods becomes deficient, and the cell fails to receive adequate quantities of energy.

Niacin deficiency leads especially to a discoloration of the skin, which becomes darkened on exposure to sunlight. Along with this, severe muscular weakness, diarrhea, and one or more psychoses often occur. This is the clinical picture called *pellagra.*

Deficiency of riboflavin is not as likely to lead to such severe difficulties as deficiency of niacin, because other types of flavoproteins that can perform almost the same functions as the protein derived from riboflavin are present in the body. Nevertheless, riboflavin deficiency, when occurring along with deficiencies of niacin or thiamine, can greatly intensify the symptoms observed in pellagra or beriberi. A common result of riboflavin deficiency alone is cracking at the angles of the mouth, called *cheilosis.*

Vitamin B_{12} and folic acid. These two substances are needed by the bone marrow to form red blood cells and are also needed in all the other tissues of the body for adequate growth. When they are lacking, the red cells released into the blood are usually considerably larger than normal, poorly formed, and very fragile. On studying the bone marrow one finds that the new red cells being formed have abnormal structures, for which reason it is believed that vitamin B_{12} and folic acid are necessary for the formation of the structural elements of the cells. However, lack of these vitamins does not affect the formation of hemoglobin.

Red cells are affected more than other cells of the body by lack of these two substances, probably because these cells are produced much more rapidly than most other cells. This is supported by the fact that cancer tissues are affected in very much the same way as red cells, for cancer cells also are produced much more rapidly than normal cells. In fact, purposeful deficiency of these two vitamins can actually be used as a means for slowing up the growth of some cancers.

Less important vitamin B compounds. For approximately the first 20 years after beriberi was discovered to be a nutritional disease, it was believed that the vitamin extract used to treat this disease contained only one single vitamin, and this was called *vitamin B.* Later, many different vitamins were discovered in the extract, and these have come to be known as the *vitamin B complex.* The vitamins included in this complex are thiamine (vitamin B_1), riboflavin (vitamin B_2), niacin, vitamin B_{12}, folic acid, *pyridoxine, pantothenic acid, biotin, inositol, choline,* and *para-aminobenzoic acid.* The functions of the latter six of these vitamins in the human being are not well known, but some of the functions of pyridoxine and pantothenic acid, the two most important of them, are the following:

PYRIDOXINE. Pyridoxine is needed for synthesis of some of the amino acids and perhaps other compounds. Dietary lack of this vitamin in lower animals can cause *dermatitis, decreased rate of growth,* development of a *fatty liver, anemia,* and different types of *mental symptoms.* The effects of lack in the human being are not too well known, because pyridoxine deficiency almost never occurs without simultaneous deficiency of other vitamins of the B complex.

PANTOTHENIC ACID. Pantothenic acid is used in the body to form a special chemical called *coenzyme A,* which catalyzes the acetylation of many substances in the cells. For example, pyruvic acid must be acetylated before it can be oxidized, and acetylcholine is formed from choline by the process of acetylation. Pantothenic acid deficiency has never been known to develop in the human being, because this substance is widespread in almost all foods of the diet. In animals, deficiency can be created artificially, and the results are retarded growth, failure of reproduction, graying of the hair, dermatitis, fatty liver, and many other effects. All these effects testify to the importance of pantothenic acid to the metabolism of the body.

Ascorbic acid (vitamin C). The major function of ascorbic acid is to maintain normal intercellular substances throughout the body, though the mechanisms by which this occurs are unknown. These include the connective tissue fibers that hold the cells together, the intercellular

cement substance between the cells, the matrix of bone, dentin of the teeth, and other substances excreted by the cells into the intercellular spaces.

Deficiency of ascorbic acid in the diet causes failure of wounds to heal because of failure to deposit new fibers and new cement substance. Also, it causes bone growth to cease and blood vessels to become so fragile that they bleed on the slightest provocation. In short, lack of vitamin C leads to loss of integrity of many of the tissues of the body, and the resulting picture, characterized especially by bleeding of the gums, splotchy hemorrhages beneath the skin, and many other internal abnormalities, is the disease called *scurvy*.

Vitamin E. In lower animals, lack of vitamin E can cause the male germinal epithelium in the testes to degenerate, causing sterility; in the female it can cause reabsorption of a fetus even after conception. For these reasons vitamin E is sometimes called the *antisterility* vitamin. However, deficiency of this vitamin in the human being has never been proved, and the precise chemical means by which vitamin E functions also has not been determined.

Severe vitamin E deficiency in animals can also cause the muscles of the hindquarters to degenerate, and because this degeneration has some of the characteristics of *muscular dystrophy* in human beings, it has been speculated that muscular dystrophy might result from vitamin E deficiency. Unfortunately, administration of vitamin E to muscular dystrophy patients has not proved to be of benefit.

Vitamin D. Vitamin D promotes calcium and phosphate absorption from the gastrointestinal tract. Without adequate quantities of vitamin D the bones become deficient in both of these substances, and in very severe vitamin D deficiency the level of ionic calcium in the blood may fall so low that muscular tetany develops. All these effects of vitamin D deficiency will be discussed in relation to calcium metabolism and parathyroid hormone in Chapter 35.

Vitamin K. Vitamin K usually is not in the diet in large quantities. Yet the normal person does not experience vitamin K deficiency because a large amount of this substance is synthesized in the colon by bacteria and is then absorbed. If the bacteria of the colon are destroyed by administration of antibiotic drugs, vitamin K deficiency usually develops within the next few days.

Vitamin K is required for the formation of prothrombin by the liver. A deficiency of vitamin K depresses blood coagulation so that there is excessive bleeding. This subject was discussed in greater detail in Chapter 11.

Mineral Requirements of the Body

Table 10 gives the amounts of different minerals and some other substances in man, and Table 11 gives the daily requirements of minerals. Some of these minerals have already been discussed in relation to other phases of the physiology of man. For instance, sodium, chloride, and calcium are major constituents of the extracellular fluid, and potassium, phosphate, and magnesium are major constituents of the intracellular fluid. These minerals are responsible for development of electrical potentials at the cell membrane and for the maintenance of proper osmotic equilibria between the extracellular and intracellular fluids. In addition, calcium and phosphate are major constituents of bone, and phosphate forms a great number of different chemicals used for a myriad of functions inside all cells, some of which were discussed earlier in this chapter. The remaining minerals that need special comment are iron, iodine, cobalt, copper, and zinc.

Iron. About two thirds of the iron of the body is in the hemoglobin in the blood; most of the remainder is stored in the liver, but to a slight extent in other tissues, in the form of *ferritin*. Ferritin can be mobilized when needed, and carried in the blood to the bone marrow, where it is used to form hemoglobin.

Iron is also present in some of the enzymes of the cells—especially in the cytochromes. Therefore, a second function of iron is to aid in the oxidation of foods in the tissue cells.

Iodine. Iodine is used by the thyroid gland to form thyroxine, a hormone that increases the rate of metabolism of the body. The functions of

Table 10. *Content in Grams of a 70 Kg. Adult Man*

Water	41,400	Mg	21
Fat	12,600	Cl	85
Protein	12,600	P	670
Carbohydrate	300	S	112
Na	63	Fe	3
K	150	I	0.014
Ca	1,160		

Table 11. *Daily Requirements of Minerals*

Na	3.0 gms.	I	250.0 μgm.
K	2.5 gms.	Mg	unknown
Cl	2.5 gms.	Co	unknown
Ca	1.0 gm.	Cu	unknown
PO_4	1.5 gm.	Mn	unknown
Fe	12.0 mgm.	Zn	unknown

thyroxine and its relation to iodine metabolism will be discussed in Chapter 33.

Copper and cobalt. Copper and cobalt both affect the formation of red blood cells. Copper, in some way yet unknown, helps to catalyze the formation of hemoglobin. Very rarely does copper deficiency exist, however, so that this is almost never a cause of hemoglobin deficiency. Cobalt is an essential element in vitamin B_{12}, and in this way is essential for maturation of red blood cells. When cobalt in other forms besides vitamin B_{12} is in the diet in excess, the bone marrow, for reasons not yet understood, produces far too many red blood cells, causing polycythemia.

Zinc. Zinc forms part of the structure of the enzyme *carbonic anhydrase,* which is present in many parts of the body, especially in the red blood cells and in the epithelium of the kidney tubules. This substance catalyzes the reaction of carbon dioxide and water to form carbonic acid and also catalyzes the same reaction in the reverse direction. It causes carbon dioxide to combine with water about 250 times as rapidly as it would otherwise. This allows the red blood cells to transport carbon dioxide far more easily than could be possible in the absence of this enzyme.

In the kidney tubules, carbonic anhydrase catalyzes the reactions that cause the secretion of hydrogen ions into the tubular fluid, thereby helping to regulate the acid-base balance of the body fluids.

In addition to the utilization of zinc in carbonic anhydrase, insulin is stored in the pancreas in the form of a zinc compound.

Fluorine. Fluorine in the diet protects against *carious* teeth. It does not make the teeth any stronger, but is believed to inactivate bacterial secretions that can cause tooth decay. Only a small trace of this substance in the drinking water of the growing child usually protects against tooth decay throughout life, which is the reason why city water supplies are now usually fluorinated.

REGULATION OF FOOD INTAKE

Food intake is regulated by the sensations of *hunger* and *appetite.* Hunger means craving for food, and the term appetite is often used in the same sense as hunger except that it usually does not imply actual discomfort as is frequently the case with hunger. Also, appetite often means a desire for specific foods instead of for food in general.

The term *satiety* means the opposite of hunger, a feeling of complete fulfillment in the quest for food. Satiety results from a filling meal.

Neural centers for regulation of food intake. Stimulation of the *lateral hypothalamus* causes an animal to eat voraciously, while stimulation of the *medial nuclei of the hypothalamus* causes complete satiety even in the presence of highly appetizing food. Therefore, we can call the lateral hypothalamus the *hunger center* or the *feeding center,* while we can call the medial hypothalamus a *satiety center.*

In addition to the hypothalamus centers, however, the conscious centers of the cerebral cortex also enter into regulation of food intake—not so much into the regulation of total quantity of intake, but into the specific choice of food, for it is in the cortex that memories of previously eaten pleasant or unpleasant foods are stored, and these memories adjust the appetite for different foods accordingly.

Long-term and short-term regulation of food intake. There are two entirely different types of food intake regulation, called respectively "long-term regulation" and "short-term regulation."

Long-term regulation means regulation of food intake in relation to the amount of nutritive stores in the body. For instance, a person who has been underfed for many weeks has intensified hunger until his normal nutritive stores have been replenished. Conversely, an animal that has been force-fed until it is greatly overweight has almost no hunger, a condition that can last for weeks until its normal body weight has returned. The precise mechanism by which the nutritive stores affect hunger is not known, though it is known that a decreased level of glucose in the body fluids, which usually goes along with decreased stores of other nutrients in the body, increases a person's hunger. Presumably, decreased quantities of amino acids and unesterified fatty acids in the body fluids cause the same effect, in this way controlling the degree of hunger.

Short-term regulation means regulation of dietary intake in relation to the amount of food that can be assimilated in a given period of time. For instance, if a person overeats, he can so overload his gastrointestinal tract that he will become sick from this alone. Therefore, during the process of eating, two principal mechanisms prevent such overeating. These are (1) "metering" of food as it passes through the mouth and (2) reflexes caused by distention of the upper gastrointestinal tract. Metering of food means that sensory receptors in the mouth and pharynx detect the amount of chewing, salivation, swallowing, and tasting and thereby quantitate the amount of food that passes through the mouth. In some way that is not understood, this information

passes to the hypothalamic feeding center to inhibit hunger. Likewise, as food fills the stomach and other regions of the upper gastrointestinal tract, visceral sensory impulses, caused mainly by distention of the gut, are transmitted to the feeding center and inhibit it. In this way, overfilling of the gastrointestinal tract is prevented until the food that is already there has had time to be digested.

Obesity. Much obesity is caused simply by overeating, which results from poor eating habits that have upset the balance of the hunger and satiety centers in the brain. For instance, many people eat three meals a day simply because of habit rather than because of hunger at the time of the meal.

Many instances of obesity also result from *inherited* imbalance between the hunger and satiety centers of the hypothalamus. For instance, when an exceedingly obese person forces himself to diet until he reduces to many pounds below his obese weight, he develops voracious hunger and, if left to his own means, will regain weight almost exactly to his original obese level. Once he reaches this level, his hunger becomes essentially the same as that of a normal person. Thereafter, he eats merely enough to maintain his weight rather than to gain still more. This is analogous to setting the thermostat in a house to a high level. In other words, the "hungerstat" is set at a higher level in these persons than in others, causing excessive eating until the person becomes very obese.

Starvation. During starvation, essentially all the stored carbohydrates in the body, which amounts to less than 300 grams of glycogen in the liver and muscles, becomes used up within the first 12 to 24 hours. Thereafter, the person exists on his stored fats and proteins. For the first 2 to 4 weeks, almost all of the energy used by the body is derived from the stored fats. But eventually even these are almost depleted, so that finally the proteins must also be used. Most tissues can give up as much as one half of their proteins before cellular death begins. Therefore, for at least another few days to a week, the body can derive its energy from proteins. But, finally, death ensues because proteins are the necessary chemical elements for performance of cellular functions. This usually occurs 4 to 7 weeks after starvation begins.

REFERENCES

Aaes-Jorgensen, E.: Essential fatty acids. *Physiol. Rev.* 41:1, 1961.

Anand, B. K.: Nervous regulation of food intake. *Physiol. Rev.* 41:677, 1961.

Boyer, P. D., Lardy, H. A., and Myrback, K.: *The Enzymes.* 2nd Ed. New York, Academic Press, 1958.

Bray, H. G., and White, K.: *Kinetics and Thermodynamics in Biochemistry.* New York, Academic Press, 1957.

Dixon, M., and Webb, E. C.: *Enzymes.* New York, Academic Press, 1958.

Drummond, G. I., and Black, E. C.: Comparative physiology: fuel of muscle metabolism. *Ann. Rev. Physiol.* 22:169, 1960.

Kleiber, M. and Rogers, T. A.: Energy metabolism. *Ann. Rev. Physiol.* 23:15, 1961.

Mayer, J., and Bullen, B.: Nutrition and athletic performance. *Physiol. Rev.* 40:369, 1960.

Mehler, A. H.: *Introduction to Enzymology.* New York, Academic Press, 1957.

Stetten, D., Jr., and Stetten, M. R.: Glycogen metabolism. *Physiol. Rev.* 40:505, 1960.

Szent-Gyorgyi, A.: *Bioenergetics.* New York, Academic Press, 1957.

BODY HEAT, AND TEMPERATURE REGULATION

32

HEAT PRODUCTION IN THE BODY

Metabolic rate. In the preceding chapter it was pointed out that essentially all of the energy released from foods eventually becomes heat. Therefore, the rate of heat production by the body, which is called the *metabolic rate,* is actually a measure of the rate at which energy is released from foods. The metabolic rate is measured in Calories, which is the same term used to express the amount of energy in foods. When a normal individual is in a very quiet state, the metabolic rate may be as low as 60 to 70 Calories per hour. On the other hand, it may be as high as 1000 to 2000 Calories per hour for a few minutes at a time, or as high as 200 to 300 Calories per hour for many hours at a time.

Factors That Affect the Metabolic Rate

Any factor that increases the rate of energy release from foods also increases the metabolic rate. Some of the more important of these are the following:

Exercise. Exercise is perhaps the most powerful stimulus for increasing the metabolic rate. When muscles contract, a tremendous quantity of adenosine triphosphate is degraded to adenosine diphosphate, and this then enhances the rate of

oxidation of the foodstuffs. During very strenuous exercise lasting only a moment or more the metabolic rate can actually increase to as much as 40 times normal.

Sympathetic stimulation and norepinephrine. When the sympathetic nervous system is stimulated, norepinephrine is released directly into the tissues by the sympathetic nerve endings. Also, large quantities of this hormone and of epinephrine are released into the blood by the adrenal medullae. These two hormones then exert a direct effect on all cells to increase their metabolic rates. The exact means by which this is accomplished is not known, but it does enhance the breakdown of glycogen into glucose and also increases the rates of some of the enzymatic reactions that promote oxidation of foods.

Strong sympathetic stimulation can increase the metabolic rate to as much as 200 per cent of normal, but the metabolic rate remains elevated for only a few minutes after the sympathetic stimulation ceases. By controlling the activity of the sympathetics, the central nervous system has a means for regulating the rates of activity of all the cells of the body, increasing their activity when this is required and decreasing their activity when the need no longer exists.

Thyroid hormone. Thyroid hormone has an action on all cells of the body similar to that of norepinephrine, except that the action of thyroid hormone continues for as long as four to eight weeks after its release from the thyroid gland, rather than for only a few minutes. Thyroid hormone secreted in very large amounts, like norepinephrine, can increase the metabolic rate to as much as 200 per cent of normal, and complete lack of secretion by the thyroid gland causes the metabolic rate to fall to as low as 50 per cent of normal. In other words, the overall span from total lack of thyroid hormone to great excess can increase the metabolic rate as much as 4-fold.

The precise mechanism by which thyroid hormone affects the cells is not known, but it is known to increase the quantities of most of the cellular enzymes, which perhaps explains its metabolic effects. The actions of thyroid hormone will be discussed at further length in Chapter 33.

Most of the other endocrine hormones, except norepinephrine and thyroxine, have only minor effects on the overall metabolic rate, though insulin from the pancreas, growth hormone from the anterior pituitary gland, and adrenocortical hormones can increase the general rate of metabolism as much as 5 to 15 per cent.

Body temperature. The higher the temperature of a chemically reactive medium, the more rapid are the chemical reactions. This effect is observed also in the chemical reactions of all cells. Each degree Centigrade increase in temperature increases the rate of any chemical reaction, in the body or out of the body, approximately 10 per cent. With very high fever the metabolic rate may be as much as twice normal because of the fever itself.

Specific dynamic action of foods. After a meal, the metabolic rate usually rises and remains elevated for the ensuing 2 to 10 hours. In general, a meal containing large quantities of fats and carbohydrates will increase the metabolic rate about 4 to 15 per cent, with the increase lasting 3 to 6 hours. After a meal containing large quantities of proteins, however, the metabolic rate may increase as much as 30 to 60 per cent, and the increase may last as long as 10 to 12 hours.

The specific dynamic action of foods is believed to be caused at least partially by the increased metabolism required for digestion, absorption, and assimilation of the foods. But in addition to this effect, proteins probably increase the metabolic rate an extra amount because of direct stimulatory effects of some of the amino acids or of some of the breakdown products of proteins.

Basal Metabolic Rate

Because of the many different factors that can affect the metabolic rate, it is extremely difficult to compare the rates of metabolism from one person to another. To have any valid comparison at all, a person's rate of energy utilization must be measured when he is in a so-called *basal* state. This means that (1) he is not exercising and has not been exercising for at least 30 minutes to an hour, (2) he is at complete mental rest so that his sympathetic nervous system is not overactive, (3) the temperature of the air is completely comfortable so that his body temperature is not too high nor too low and also so that his sympathetic nervous system is not unduly stimulated, (4) he has not eaten any food within the last 12 hours that could cause a specific dynamic action on the metabolic rate, and (5) his body temperature is normal to avoid the effect of fever on metabolism.

Under *basal conditions* most of the factors that affect the metabolic rate are controlled, but the basal state does not remove the effect of the constantly secreted thyroid hormone. Therefore, the *basal metabolic rate* in essence is determined by two major factors: first, the inherent rates of chemical reactions of the cells, and, second, the amount of thyroid hormone acting on the cells. Because the inherent rate of cellular activity is relatively constant from one person to another,

an abnormal basal metabolic rate is usually caused by either excessive or deficient secretion of thyroid hormone. For this reason basal metabolic rates are often measured to assess the degree of thyroid activity.

Direct method for measuring basal metabolic rate. Since the basal metabolic rate is actually a measure of the amount of heat produced by the body under basal conditions, it can be determined by measuring the heat given off from the body in a known period of time. To do this, human subjects are placed in a large chamber called a *human calorimeter.* This chamber is cooled by water flowing through a radiator system. The heat given off from the body is picked up by the cooling system and then measured by appropriate physical apparatus. This is called the *direct* method for measuring the basal metabolic rate because it measures the heat output directly. Obviously, it is very cumbersome, but it has been a very important tool in experimental studies.

Indirect method for measuring basal metabolic rate. An indirect method for measuring the basal metabolic rate is based on the amount of oxygen burned by the body in a given period of time. From this the rate of energy release can be calculated. The amount of energy released when 1 liter of oxygen burns with carbohydrates is 5.05 Calories. When 1 liter of oxygen burns with fats the amount released is 4.70 Calories. And when 1 liter of oxygen burns with proteins the amount is 4.60 Calories. It is obvious from these figures that every time 1 liter of oxygen is burned in the metabolic fires of the body, essentially the same amount of energy is released regardless of which one of the three different foods is being used for energy. Therefore, it is reasonable to use an approximate average of these values, 4.825 Calories, as the amount of energy released in the body every time 1 liter of oxygen is burned. When calculating the basal metabolic rate in this way, the value obtained is never more than 4 per cent in error, even though a great excess of one type of food or the other might be used for metabolism momentarily, and 4 per cent is far less than the error of measurement anyway. Therefore, to determine the amount of energy being generated in the body, one needs only to determine the amount of oxygen being used. This is done with a respirometer.

THE RESPIROMETER. Figure 309 shows a respirometer. The subject places a mouthpiece in his mouth and breathes in and out of an inverted can that rides up and down in a water bath. The respirometer contains oxygen that is breathed back and forth into the lungs. The oxygen is gradually absorbed into the blood, and in its place carbon dioxide is expired. Soda lime in the respirometer reacts chemically with the carbon dioxide to remove it from the respiratory gases. The net result, therefore, is a continual loss of oxygen from the respirometer; this loss causes the inverted can to sink deeper and deeper into the water. The recording pen also falls lower and lower, giving a record of the rate of oxygen utilization.

Method for expressing the basal metabolic rate. The basal metabolic rate is generally expressed in terms of Calories per square meter of body surface area per hour. The reason for expressing it in terms of body surface area is to allow comparisons between individuals of different sizes. Experimental studies have shown that the basal metabolic rate varies from one normal person to another approximately in proportion to the body surface area and not in proportion

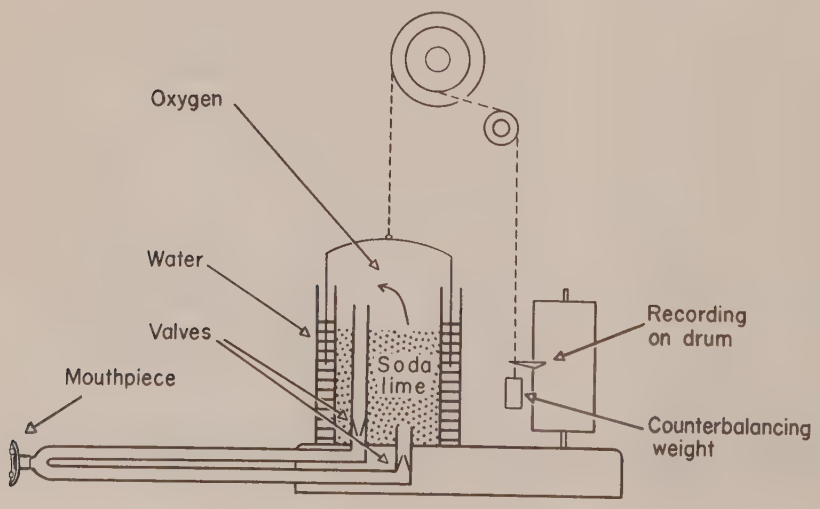


Figure 309. A respirometer for measuring the rate of oxygen utilization.

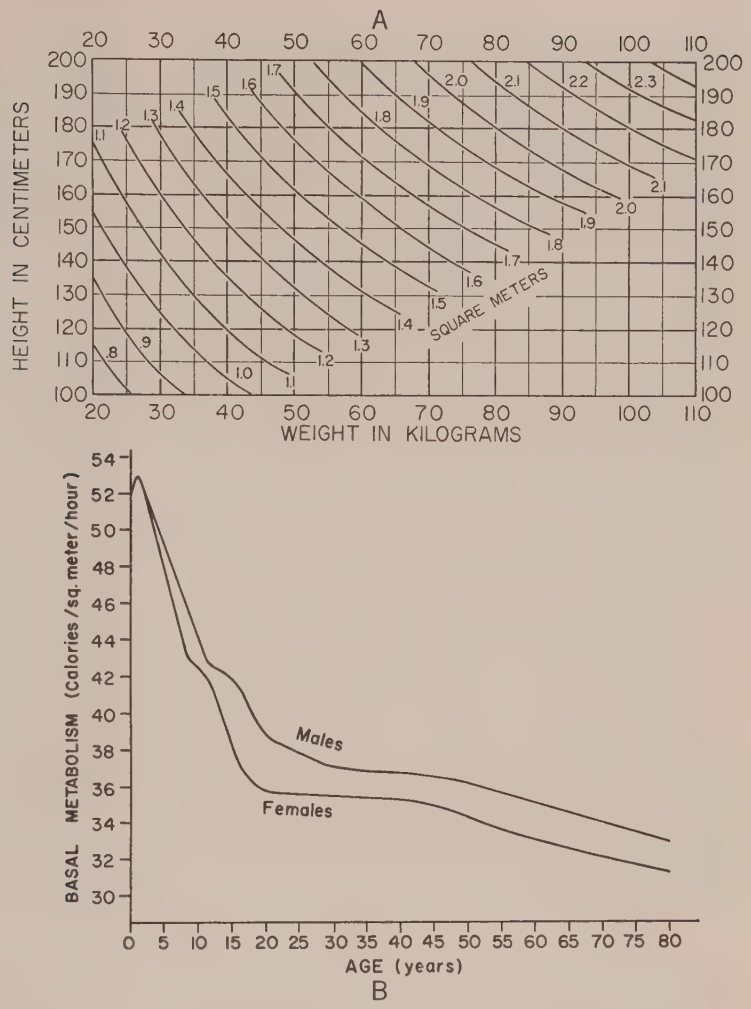


Figure 310. (A) Diagram for determining the body surface area when the weight and height are known. (From DuBois: *Basal Metabolism in Health and Disease.* Lea & Febiger.) (B) Normal basal metabolic rates for males and females at different ages.

to the weight. Thus, someone who weighs 200 pounds does not have a basal metabolic rate twice that of someone weighing 100 pounds, but only about 30 per cent greater. His surface area also is about 30 per cent greater.

A representative calculation of basal metabolic rate might be the following: A person is found to use 1.8 liters of oxygen in six minutes. Therefore, in one hour he uses 18 liters of oxygen, and his total basal metabolic rate for this period would be 18 times 4.825, or 86.85 Calories per hour. To express this in Calories per square meter per hour one uses the chart in Figure 310A, which shows the square meters of surface area in relation to weight and height. If the person weighs 70 kg. and is 180 cm. tall, his surface area is 1.9 square meters. Therefore, his basal metabolic rate is 86.85 Calories per hour divided by 1.9, or 45.7 Calories per square meter per hour.

The basal metabolic rate is often also expressed in per cent of normal. To do this one refers to the chart in Figure 310B, which shows the normal basal metabolic rate for males and females at different ages. If this person is a boy aged 18, his normal basal metabolic rate is 40 Calories per square meter per hour. However, his actual basal metabolic rate is 5.7 Calories greater than the normal value, or 14 per cent greater than normal. His basal metabolic rate, therefore, is stated to be +14. If his basal metabolic rate had been 25 per cent less than normal, it would have been expressed as −25, rather than plus.

When the thyroid gland is secreting an extreme quantity of thyroxine, the basal metabolic rate can at times go as high as +100. On the other hand, when the thyroid gland is secreting almost no thyroxine, the basal metabolic rate falls to as low as −40 to −50.

Major Loci of Heat Production in the Body

Though all the tissues of the body produce heat, those that have rapid chemical reactions

produce the major amounts. In the resting state, the liver, heart, brain, and most of the endocrine glands produce large amounts of heat. This causes their temperature to be a degree or so higher than that of most of the other tissues.

During rest the amount of heat produced by each skeletal muscle is not very great, but, because half of the entire mass of the body is composed of muscles, heat production of all the skeletal muscles together nevertheless accounts for about 40 per cent of all the body's heat production even at rest. A slight increase or decrease in the degree of muscular tone can have considerable effect on the amount of heat produced. During severe exercise the amount of heat produced by the muscles can rise to as great as 20 times that produced by all the remaining tissues together. For this reason, changes in the degree of muscular activity constitute one of the most important means by which the body regulates its temperature. This is discussed in detail later in this chapter.

HEAT LOSS FROM THE BODY

The heat produced in the body must be removed continually or otherwise the body temperature would continue rising indefinitely. The body loses heat to its surroundings in three ways: radiation, conduction, and evaporation.

Heat Loss by Radiation

About 60 per cent of the heat loss from a nude person sitting in a room at 70°F. is by radiation, as shown in Figure 311. Heat loss in this manner is based on the principle that objects near each other are always radiating heat toward one another. The human being radiates heat toward the walls, and the walls radiate heat toward him. However, since his body temperature is usually greater than that of the walls, he radiates more heat to the walls than he receives from them.

The principle of radiation is used in radiant heating systems in commercial buildings and in many homes. The floors, ceilings, and walls of the rooms are heated to temperatures between 70 and 85° F. These are temperatures approximately comfortable for the human being, because they permit him to radiate heat at a rate that allows his body temperature to remain constant. A person is often comfortable in a radiantly heated room even though the air temperature may be in the 50's. This fact illustrates very forcefully how important radiation can be as a means of heat exchange.

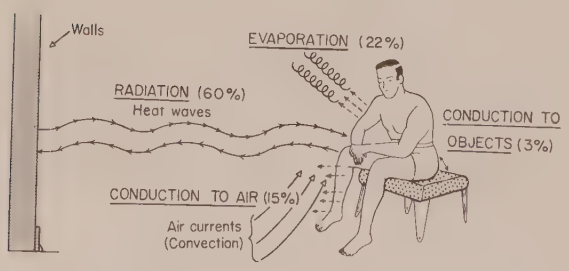


Figure 311. Methods by which heat is lost from the nude body. The percentage values are those for heat loss by each method when the person is in a room at 70° F.

Heat Loss by Conduction

Approximately 18 per cent of the heat lost by the nude person in Figure 311 leaves his body by conduction, 15 per cent by conduction to the air and 3 per cent to the floor and stool. These values are only average, for the colder the air and adjacent solid objects, the greater is the conduction of heat into them.

Effect of convection currents on loss of heat by conduction. Even though the temperature of the air remains constant at 70° F., the rate of conduction of heat into the air depends on how much the air is moving. If it is flowing past the body, every time the air adjacent to the body becomes warmed it is carried away to be replaced by cooler air. The more rapidly the air moves, the greater is the amount of heat conducted from the body. For this reason it is sometimes said that large quantities of heat are lost from the body by convection. This is not actually true; the heat is still being lost by conduction, though the convection currents carry the heated air away.

Heat Loss by Evaporation

A small amount of extracellular fluid continually diffuses through the skin and evaporates, and evaporation of each gram of water removes approximately one-half Calorie of heat from the body. Evaporation of only 150 ml. of water each hour would remove all the heat produced under basal conditions. This shows how important evaporation can be as a cooling mechanism.

The sweat mechanism. In addition to the continual diffusion of water through the skin, the sweat glands produce large quantities of sweat when the body is exposed to extreme heat. The

sweat obviously increases the amount of heat that can be lost by evaporation. Under extreme conditions about one gallon of sweat can be secreted each hour, which can remove as much as 2000 Calories of heat from the body, or about 32 times the basal level of heat production.

Necessity for evaporation in tropical climates. Evaporation as a means of heat loss is extremely important in tropical climates, for when the temperature of the air and of the surroundings rises above the temperature of the body, heat cannot be lost by either radiation or conduction. Instead, the body gains heat by these means. However, the evaporation of sweat can still keep the surface of the body at a temperature lower than that of the surroundings. Therefore, one's only means for maintaining normal body temperature in the tropics is by sweating. Those few persons who are born without sweat glands must forever live in temperate or cold climates.

Effect of convection on evaporation. Air currents have much the same effect on the removal of heat by evaporation as by conduction, for water evaporating from the skin quickly saturates the air immediately adjacent to the skin. If this air does not move away rapidly, the evaporation process will cease. However, if new air continually replaces the old, the air next to the skin never becomes saturated with moisture, and evaporation can continue unabated. This explains why in hot climates a fan is an important means for keeping cool; it also explains why one is usually much cooler outdoors under the shade of a tree, where the air is moving, than he is inside a house where the air is not moving, even though the air temperature may be the same.

Effect of Clothing on Heat Loss

Clothing is a barrier to the transfer of heat from the body to the surroundings. Heat radiated or conducted from the skin is absorbed by the inner surface of the clothing, but before it can be radiated or conducted to the surroundings it must pass through the mesh of the cloth. The inside of the clothing becomes warm in comparison with the outside; the inside warmth decreases the rate of conduction of heat from the body to the clothing and even radiates much of the heat back to the body.

The insulation properties of most clothing result mainly from air entrapped within the clothing mesh and not from the actual material comprising the clothing. A furred animal utilizes this principle in the winter time, for his hair grows long and stands on end, entrapping large amounts of air that become warm and act as a body insulator.

When clothing becomes wet, the spaces are filled with water, which conducts heat many hundred times as rapidly as does air. As a result, the clothing is no longer an adequate heat insulator; instead, heat is conducted almost as if the clothing did not exist. Therefore, wet clothing has almost no value in keeping the body warm in cold climates, and one of the most important lessons to be learned for arctic survival is always to keep the clothing dry whatever the circumstances.

Effect of clothing in the tropics. In the tropics, the effect of clothing on conduction and radiation usually is not important, because the temperature of the surroundings is close to that of the body. Instead, clothing must be designed for two other purposes: to protect one from direct exposure to the sun's heat rays and to allow maximum evaporation of sweat. Almost any clothing that blocks light rays also blocks the heat rays of the sun, but different types of clothing have vastly different effects on evaporation. Those types of clothing that rapidly absorb water, such as the cottons and the linens, usually do not depress the evaporative cooling of the body, because sweat is absorbed into the clothing and evaporated from its surface in the same manner as from the body. This is not true of wool and plastic materials, for they are not absorptive enough to allow satisfactory evaporative cooling.

THE BODY TEMPERATURE

The body temperature is determined by the balance between heat production and heat loss, as shown by Figure 312. If these two are exactly equal, the body temperature neither rises nor falls. When heat production is greater than heat loss, the body temperature rises; conversely, when heat loss is greater than heat production, the body temperature falls. Appropriate regulatory systems are always at work in the body to keep heat production and heat loss approximately equal, thereby maintaining a normal body temperature. These regulatory mechanisms are discussed as follows.

Normal values of body temperature. The normal body temperature varies from one person to another no more than a degree or so. However, when a person is exposed to extremely cold or hot weather, his overall body temperature may vary as much as 1° F. Also, when he experiences extreme emotions that cause excessive stimulation of the sympathetic nervous system, the amount of

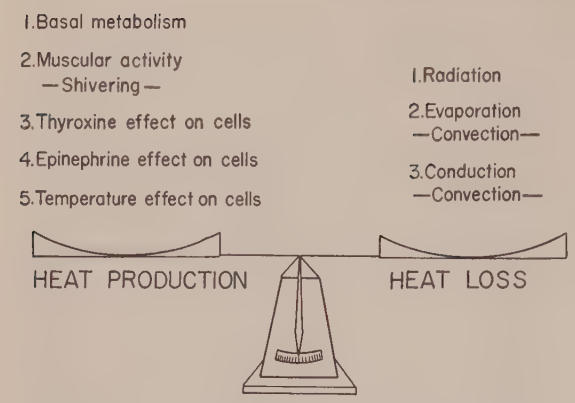


Figure 312. Balance between heat production and heat loss, illustrating that the body temperature remains constant as long as these two are equal.

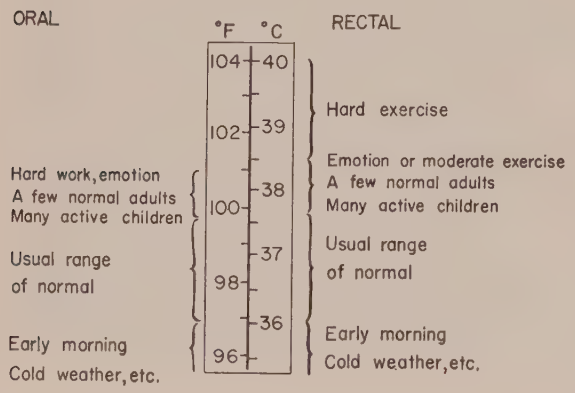


Figure 313. Ranges of body temperature under different normal conditions. (From DuBois: *Fever and the Regulation of Body Temperature.* Charles C Thomas.)

heat production may become great enough to raise the body temperature a degree or so. And, finally, extremely hard exercise can sometimes increase the body temperature as much as 5 to 6° F., though usually within 10 to 20 minutes after the exercise is over the temperature will have fallen back to that of the basal state.

Figure 313 depicts the normal temperature under different conditions as measured in the mouth and also as measured rectally. It is evident from this figure that the rectal temperatures are usually slightly higher than the oral temperatures. The average normal oral temperature is usually considered to be between 98.0 and 98.6° F., while the average normal rectal temperature is generally considered to be about 1° F. higher. The reason for the higher rectal temperature is that the mouth is continually cooled by the facial surfaces and by evaporation in the mouth and nose.

REGULATION OF BODY TEMPERATURE

Hypothalamic Regulation of Temperature

Nervous centers located in the hypothalamus are capable of keeping the body temperature within the normal range essentially all the time. The hypothalamus has two major divisions for the regulation of body temperature: the anterior *heat losing center* which, when stimulated, reduces the body heat and the posterior *heat promoting center* which, when stimulated, increases the body heat. These areas are shown in Figure 314. The anterior center corresponds mainly with the parasympathetic centers, while the posterior center is principally sympathetic in nature.

Hypothalamic mechanisms for increasing the body temperature. Whenever blood colder than normal passes into the hypothalamic regions of the brain, the posterior heat promoting center is strongly activated. Why cold blood causes this reaction is not yet known, but stimulation of this center automatically activates several different mechanisms to increase the body heat. These are the following:

VASOCONSTRICTION. Stimulation of the heat promoting center acts through the sympathetics to cause the blood vessels of the skin to constrict very powerfully. This decreases the flow of warm blood from the internal structures to the skin, decreasing the transfer of heat to the body surface from those organs producing most of the heat. Very little heat diffuses directly from the internal structures to the body surface by means other than through the blood because of the fat beneath the skin which is a very adequate heat insulator.

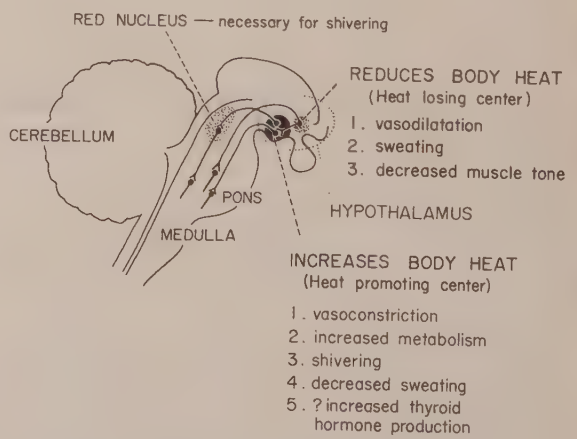


Figure 314. The hypothalamic and brain stem mechanisms for regulating body temperature.

The temperature of the skin falls to approach the temperature of the surroundings; this causes heat loss to be greatly diminished, allowing the internal body temperature to rise.

INCREASED METABOLISM. Stimulation of the sympathetic nerves releases norepinephrine throughout the body tissues. The norepinephrine in turn increases the rates of metabolism of all cells, thereby increasing heat production. This, too, tends to increase the body temperature.

SHIVERING. Stimulation of the heat promoting center increases the degree of wakefulness, and also causes transmission of large numbers of impulses into the bulboreticular formation and red nucleus of the hindbrain. Impulses passing through these regions increase the tone of all the muscles, which in turn increases the amount of heat produced by the muscles. When the tone becomes extremely great the stretch reflex begins to oscillate; that is, a slight movement stretches a muscle which elicits the stretch reflex causing the muscle to contract. The contraction stretches the antagonistic muscle which also develops a stretch reflex. Its contraction then stretches the first muscle, and a continuing shaking develops. All this muscular activity can increase the rate of heat production by as much as several hundred per cent. Consequently, when the body is exposed to extreme cold, shivering is a very powerful force to maintain normal body temperature.

INCREASED THYROID HORMONE PRODUCTION. Stimulation of the heat promoting center does not have any immediate effect on the production of thyroid hormone, but if the body is exposed to cold for several weeks, as at the beginning of winter, the thyroid gland begins to produce increased quantities of thyroid hormone. It is believed that this may be caused by formation of a neurosecretory hormone in the hypothalamus (see Chapter 33) that passes in the blood to the anterior pituitary gland to increase its production of thyrotropic hormone that in turn excites the thyroid gland. The increase in thyroid hormone production over a period of weeks increases the rate of heat production 10 to 20 per cent and allows one to withstand the prolonged cold of wintertime.

PILO-ERECTION. Pilo-erection means that hairs stand on end. This occurs when the sympathetic centers are stimulated, for the sympathetic nerves excite small pilo-erector muscles located at the bases of the hairs. In the human being this mechanism does not protect against heat loss because of the scarcity of hairs, but in lower animals pilo-erection entraps large quantities of air in the zone adjacent to the body and provides increased insulation against cold.

Hypothalamic mechanisms for decreasing the body temperature. When the temperature of the hypothalamus rises too high, the anterior hypothalamic heat losing center becomes stimulated, and because of reciprocal innervation between this center and the heat promoting center, the latter is inhibited. As a result, all the mechanisms of the heat promoting center that tend to increase the body temperature become inoperative. For example, instead of vasoconstriction, the blood vessels to the skin dilate, allowing the skin to become very warm so that heat can be lost rapidly. Also, the increased metabolism caused by sympathetic stimulation ceases, muscular tone greatly decreases, and the production of thyroid hormone diminishes gradually. The reversal of all of these effects allows the rate of heat loss to increase and the rate of heat production to decrease, thus making the body temperature fall.

In addition to reversing the effects of sympathetic stimulation, stimulation of the heat losing center causes two effects of its own to decrease the temperature. These are sweating and panting.

SWEATING. If the reversal of the heat promoting effects is not sufficient to bring the body temperature back to normal, the anterior hypothalamus initiates the sweating process. As much as a gallon of sweat can be poured onto the surface of the skin in an hour, and under favorable conditions a large proportion of this will evaporate. By this means, a refrigeration mechanism is initiated to reduce the body temperature when it tends to rise too high.

PANTING. In lower animals, though not in the human being, excessive stimulation of the hypothalamus by heat initiates a neurogenic mechanism in the pons of the hindbrain to cause panting. The animal breathes very rapidly but very shallowly, so that a tremendous quantity of air passes into and out of the respiratory passageways. Evaporation of water from the respiratory passageways is greatly increased, allowing a major amount of heat loss. Many lower animals—dogs, for instance—do not have well developed sweat mechanisms, and panting is often their only means of regulating body temperature in hot climates. Because of the shallowness of breathing during panting, most of the air entering the alveoli is dead space air, so that the alveolar ventilation of the animal remains essentially normal. This prevents overventilation despite the great turnover of air in the respiratory passageways.

Summary of hypothalamic functions in body temperature regulation. The hypothalamus, when exposed to a temperature above normal, decreases heat production and increases heat loss. Conversely, when exposed to a temperature below normal, it increases heat production and decreases heat loss. The hypothalamus acts as a

thermostat, maintaining the internal body temperature usually within one-half degree of the normal average value. The efficiency of the hypothalamus as a thermostat is indicated by the fact that the nude body can be exposed for many hours to dry air temperatures as low as 50° F. or as high as 170° F. without changing the internal body temperature more than 1 to 2 degrees.

Effect of Peripheral Reflexes on Body Heat

In addition to the hypothalamic mechanism for regulating body temperature, a few peripheral reflexes also aid in its regulation to a much less extent. Each area of the skin is supplied by spinal cord reflexes that dilate the blood vessels whenever the area is exposed to heat and constrict the vessels when the skin is exposed to cold. This explains why a hand placed in hot water soon becomes red, or a hand placed in cold water soon becomes blanched and white. These reflexes obviously aid in promoting heat loss from the skin in hot surroundings and in conserving heat in cold surroundings.

Despite this aid of the cord reflexes, a person whose spinal cord has been transected in the neck region, and who therefore has lost connections between the hypothalamus and the cord, is almost completely unable to regulate his body temperature. His hypothalamus is unable to transmit impulses into the sympathetics or into the fibers that control the sweat glands, and the bulboreticular formation and red nucleus cannot initiate the shivering mechanism. To keep his body temperature normal, the temperature of his surroundings must be maintained at a precise level and adjusted often to his body demands. Even a few minutes of exposure to the hot sun can increase his body temperature to almost lethal levels.

FEVER

Fever means a body temperature that is elevated beyond the normal range. It occurs in many disease states and, therefore, is of extreme importance in assessing the severity of a person's affliction. The most frequent cause of fever is severe bacterial or viral infection such as occurs in pneumonia, typhoid fever, tuberculosis, diphtheria, measles, yellow fever, mumps, poliomyelitis, and virus pneumonia. A less frequent cause of fever is the destruction of body tissues by some means other than infection. For instance, a person having a severe heart attack usually develops considerable fever, and a patient exposed to destruction of his tissues by x-ray or nuclear radiation may have fever for the ensuing few days.

Fever Caused by the Resetting of the Hypothalamic Thermostat by Proteins

Fever is usually caused by abnormal proteins released into the body fluids during disease processes. These proteins have a direct effect on the hypothalamic thermostat to reset its normal operating range at a higher temperature level. For example, only one thousandth of a gram of protein derived from the bodies of typhoid bacilli can reset the level of the thermostat from normal up to as high as 110° F. When this happens, all the heat promoting mechanisms for increasing the body temperature become active and remain active until the body temperature reaches 110° F. Then the mechanisms for heat loss and heat production become equalized again, and the temperature continues to be regulated at this elevated level as long as the abnormal protein is present. In other words, the body temperature is still regulated during fever, but it is regulated at a level considerably above normal because of resetting of the thermostat.

Chills and sweating in feverish conditions. Figure 315 shows the effect of disease on the course of the body temperature. At the beginning of this chart, the oral temperature is 98.6° F., but after a few hours the disease suddenly sets the thermostat to a level of 103° F. During the ensuing hours all the heat promoting mechanisms for increasing body temperature operate at full force. These include vasoconstriction, increased metabolism caused by norepinephrine secretion, and shivering. Even when the body temperature reaches 101° F., it still has not reached the setting

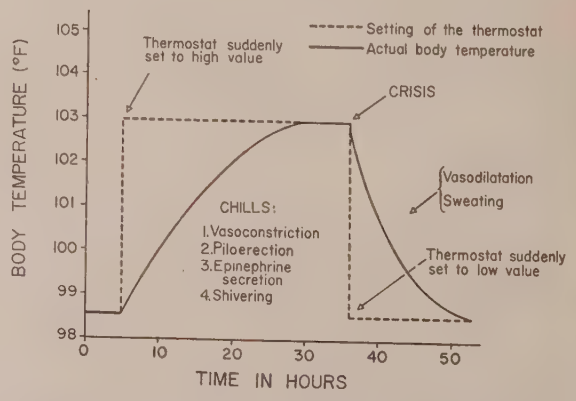


Figure 315. Fever caused by a disease process; development of chills when the temperature is rising and development of sweating when the temperature is falling.

of the thermostat, and the heat promoting phenomena continue to take place. Therefore, even though the temperature is high, the skin remains cold, and shivering occurs. This is called a *chill;* when someone is having a chill it is quite certain that his body temperature is actively rising.

After a number of hours the body temperature reaches the setting of the thermostat, and the chills disappear, but the temperature remains regulated at the high thermostatic setting until some factor breaks the disease process. In Figure 315 the disease process is corrected after another few hours, and the thermostat is suddenly set back to its normal value of 98.6° F. However, the body temperature is still high. This causes the heat losing mechanisms, including especially vasodilatation and sweating, to decrease body temperature. The person's skin suddenly becomes warm, and he begins to sweat, which effect is known as the *crisis.* When this happens one knows that the patient's temperature is beginning to fall. In the days before the miracle drugs, the sulfonamides and the antibiotics, many persons with bacterial diseases died, and the duty of the doctor was mainly to keep the sick person comfortable until he either died or a crisis appeared. If the crisis did appear, the doctor left the sickroom and announced to the family that all was now going to be well. The family in turn revered the doctor for his miraculous achievement. Unfortunately for the reputation of the present-day doctor, he too frequently cures the patient long before a crisis can develop, and any chance for good will is often lost.

Possible value of fever in disease. One wonders what the value of fever in disease might be. At present no clear-cut answer is available, but suggested values are: (1) Many bacterial and viral agents do not survive as well at high temperatures as at normal body temperatures. Therefore, an elevation in temperature might well be a means for combating these infections. For example, high temperature has been shown to be especially lethal to gonococcal and syphilitic bacteria. (2) Because high rates of chemical reactions occur in the cells at high temperatures, it is possible that these increased rates allow the cells to repair the damage much more rapidly than they could otherwise.

Effect of Temperature Itself on the Body

When the body temperature rises above 106° F. it becomes very difficult, and often impossible, for the temperature regulatory mechanisms to return the body temperature to normal again. The main reason for this is that at these high temperatures the rates of cellular metabolism become so greatly increased because of the temperature itself that often no amount of regulation can overcome this very rapid rate of heat production. As a result, the body temperature rises still higher, the rate of heat production rises higher, and a vicious cycle develops, causing the body temperature to keep on rising.

When the body temperature rises above 108° F. the metabolic rates of the cells become so great that the cells begin to "burn themselves out," and when the body temperature rises to 112 to 114° F. death almost always ensues simply because of the heat itself. The most damaging effect of a very high body temperature occurs in the neuronal cells of the brain, sometimes causing permanent destruction of these even though the person recovers.

When the body temperature falls far below normal, heat regulation again becomes impaired, but for opposite reasons. The low body temperature causes such slow rates of chemical reactions in the cells that no amount of regulation can bring the rate of heat production up high enough to return the body temperature to normal. The cold temperature decreases the rate of metabolism, and this allows the temperature to fall still lower, which decreases the rate of metabolism more, again creating a vicious cycle until the body temperature falls so low that the person dies. Usually death occurs when the body temperature reaches 70 to 75° F.

Body temperatures down to about 85° F. do not cause damage to the body, though the bodily functions become so greatly slowed that the person remains in a state of suspended animation until he is rewarmed.

REFERENCES

Atkins, E.: Pathogenesis of fever. *Physiol. Rev.* 40:580, 1960.

Carlson, L. D.: Temperature. *Ann. Rev. Physiol.* 24:85, 1962.

DuBois, E. F.: *Basal Metabolism in Health and Disease.* 3rd Ed. Philadelphia, Lea & Febiger, 1936.

DuBois, E. F.: *Fever and the Regulation of Body Temperature.* Springfield, Ill., Charles C Thomas, 1948.

Hardy, J. D.: Physiology of temperature regulation. *Physiol. Rev.* 41:521, 1961.

Kleiber, M.: Body size and metabolic rate. *Physiol. Rev.* 27:511, 1947.

Lyman, C. P., and Chatfield, P. O.: Physiology of hibernation in mammals. *Physiol. Rev.* 35:403, 1955.

Newburgh, L. H.: *Physiology of Heat Regulation and the Science of Clothing.* Philadelphia, W. B. Saunders Company, 1949.

Smith, R. E., and Hoijer, D. J.: Metabolism and cellular function in cold acclimation. *Physiol. Rev.* 42:60, 1962.

Vernberg, F. J.: Comparative physiology: Latitudinal effects on physiological properties of animal populations. *Ann. Rev. Physiol.* 24:517, 1962.

Section Nine

ENDOCRINOLOGY AND REPRODUCTION



INTRODUCTION TO ENDOCRINOLOGY: THE HYPOPHYSEAL HORMONES AND THYROXINE

33

The Nature of Hormones and Their Function

A hormone is a chemical substance elaborated by one part of the body that controls or helps control some function elsewhere in the body. In general, hormones are divided into *local hormones,* which affect cells in the immediate vicinity of the organ secreting the hormone, and *general hormones,* which are emptied into the blood by endocrine glands and flow throughout the entire circulation to affect cells in far distant parts of the body. Some general hormones affect all cells; others affect only certain specific cells. For example, growth hormone secreted by the pituitary gland and thyroxine secreted by the thyroid gland affect all cells of the body. On the other hand, the pituitary gland produces three gonadotropic hormones that affect only the sex organs, even though these hormones are secreted into the general circulation.

Regulatory functions of hormones. Hormones make it possible for one part of the body to control the functions of other parts. Some examples of the regulatory functions of local hormones are: (1) Acetylcholine secreted by skeletal nerve endings diffuses into the muscle fiber and

excites it. (2) Gastrointestinal hormones secreted by the intestinal mucosa pass to adjacent regions of the gastrointestinal tract to control the secretion of digestive juices and to control gastrointestinal motility. (3) Histamine secreted by damaged cells anywhere in the body causes increased permeability in the local capillaries, resulting in local edema.

Some examples of the regulatory functions of the general hormones are: (1) The adenohypophyseal gland secretes six different hormones that regulate respectively the rate of growth of all tissues of the body, the rate of secretion of thyroid hormone by the thyroid gland, the rate of secretion of adrenocortical hormone by the adrenal gland, and the rates of secretion of several different sexual hormones. (2) Thyroxine secreted by the thyroid gland controls the rate of metabolism of all cells. (3) Adrenal medullary hormones, which have been discussed in connection with the sympathetic nervous system, cause the same reactions throughout the body as stimulation of the general sympathetic nervous system. (4) Hormones secreted by the adrenal cortex regulate reabsorption of sodium from the kidneys, and also regulate some aspects of carbohydrate, fat, and protein metabolism. (5) Insulin secreted by the pancreas regulates the utilization of glucose throughout the body. (6) Hormones secreted by the testes control the sexual and reproductive functions of the male. (7) Hormones secreted by the ovaries control the sexual and reproductive functions of the female. (8) Hormones secreted by the parathyroid glands regulate the concentrations of calcium and phosphate in the body fluids.

The control functions of the local hormones have been discussed at different points throughout this text, and various aspects of the general hormones have also been mentioned. In the present chapter and the following two chapters the functions of the general hormones are considered in detail. The study of these and their functions is called *endocrinology,* and the glands that secrete the general hormones are called *endocrine glands.* The word *endocrine* means secretion to the interior of the body as opposed to secretion to the exterior as is true of the intestinal glands, the sweat glands, and others, all of which are called *exocrine glands.*

THE HYPOPHYSEAL HORMONES

The *hypophysis,* known also as the pituitary gland, is about the size of the tip of the little finger, and it lies in a small bony cavity beneath the base of the brain, as shown in Figure 316. It is

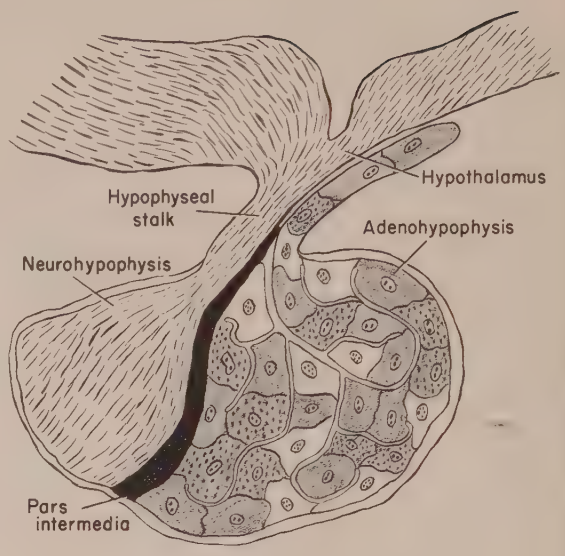


Figure 316. The hypophysis.

divided into two completely separate parts, the *neurohypophysis,* known also as the posterior pituitary gland, which is connected by a stalk with the hypothalamus of the brain, and the *adenohypophysis,* called also the anterior pituitary gland, which lies anterior to the neurohypophysis and has functions in no way related to those of the neurohypophysis.

The Adenohypophyseal Hormones

The adenohypophysis secretes at least six different known hormones, and various research workers have postulated as many as thirty more that might be produced by this gland. However, only the six have stood the test of time, for which reason it is believed that most of the other postulated hormones are really combinations of these six, which are: *growth hormone, thyrotropin, corticotropin, follicle stimulating hormone, luteinizing hormone,* and *luteotropic hormone* (Fig. 317). The last three of these are called *gonadotropic hormones* because they regulate the functions of the sex glands.

The adenohypophysis is composed of three different types of cells: *chromophobe cells, basophilic cells,* and *acidophilic cells.* Most of the hormones of the anterior pituitary gland are believed to be secreted by the basophils and acidophils, and the chromophobes seem to be only developmental stages of the other two types of cells.

Growth hormone. Growth hormone is secreted by the adenohypophysis in large quantities from the time of birth until adolescence. At this

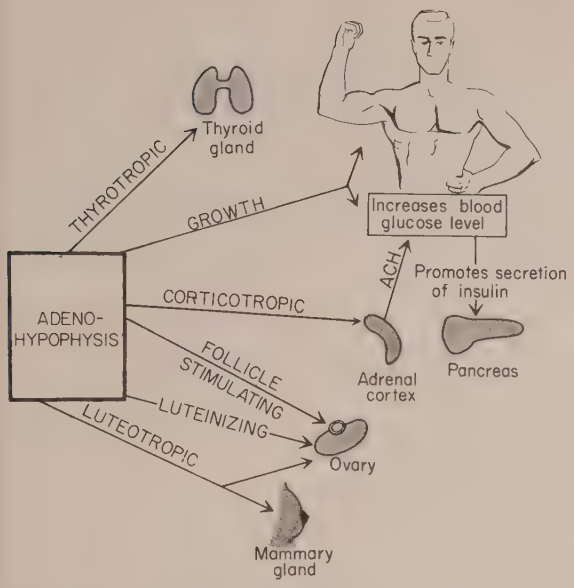


Figure 317. Adenohypophyseal hormones and their functions.

time the production of growth hormone diminishes but does not stop completely. The function of growth hormone is to promote development and enlargement of all bodily tissues. It increases the sizes of all bodily cells and also their number. Consequently, each organ becomes larger under the influence of growth hormone. The bones enlarge and lengthen, the skin thickens, and soft tissues, including the heart, liver, tongue, and all the other internal organs, increase in their size. In other words, growth hormone is exactly what its name implies; it causes the person to grow.

Growth hormone promotes growth by increasing the rates of synthesis of proteins and of other cellular elements, thereby increasing the sizes of the cellular structures. Also, the metabolic rates of the cells are increased about 15 per cent when normal amounts of growth hormone are secreted.

DWARFS. A person whose adenohypophysis fails to secrete growth hormones fails to grow. This causes the so-called "pituitary" type of dwarf, though most dwarfs have small stature because of their heredity rather than because of adenohypophyseal failure. The pituitary type of dwarf remains childlike in all physical respects. The organs as well as the bones fail to grow, and the degree of sexual development remains that of a young child even after the dwarf has reached the age of an adult. There are many degrees of pituitary dwarfism, for the secretion of growth hormone is often only diminished rather than completely lacking. If the secretion is zero, however, the person may never grow to more than twice the height of a newborn baby.

GIANTS. Secretion of excess growth hormone, if it occurs before adolescence, can cause a giant. After adolescence the growing parts of most bones have "fused" and are no longer capable of growing longer regardless of the amount of growth hormone available. Giantism usually results from a tumor of the acidophilic cells of the pituitary gland, this tumor secreting tremendous quantities of growth hormone. All parts of the body enlarge, including the skeleton, muscles, internal structures, and the skin.

ACROMEGALY. If an acidophilic tumor develops after adolescence, the excessive secretion of growth hormone cannot cause increased height; nevertheless, it can still cause increased size of the soft tissues. It can also cause thickening of the bones, and a few of the bones, the so-called "membranous bones," can even continue to grow longer. As a result, disproportionate growth occurs in different parts of the body, resulting in the condition called *acromegaly.* In acromegaly the lower jaw, which is formed of membranous bone, grows excessively long, causing the chin sometimes to protrude a half inch or more beyond the remainder of the face. The nose and lips also enlarge, and the internal organs such as the tongue, the liver, and the various glands enlarge. The bones of the feet and hands, unlike most bones of the body, continue to grow so that the hands and feet become very large. In short, an acromegalic person is the same as a giant, except that failure of most of his bones to grow causes his height to be normal while other features of the body exhibit excessive growth.

Thyrotropin. Thyrotropin, also known as thyrotropic hormone, controls the amount of hormone secreted by the thyroid gland. It increases the number of thyroid cells, the size of these cells, and also their rate of thyroxine production. When the adenohypophysis fails to secrete thyrotropin, the thyroid gland becomes so incapacitated that it secretes no measurable amount of hormone. In other words, the thyroid gland is completely controlled by thyrotropin. Further relationships of the thyroid gland to thyrotropin are discussed later in the chapter in connection with thyroxine.

Corticotropin. A third hormone secreted by the adenohypophysis, *corticotropin,* also known as adrenocorticotropic hormone, controls secretion of adrenocortical hormones by the adrenal cortices in much the same manner as thyrotropin controls secretion by the thyroid gland. Corticotropin increases both the number of cells in the adrenal cortex and their degree of activity, resulting in an increased output of adrenocortical hormones. The relationship of corticotropin to the different adrenocortical hormones is discussed in the following chapter.

The gonadotropic hormones. The functions

of the three gonadotropic hormones are summarized briefly in the following paragraphs, but their relationship to sexual function will be presented in Chapter 36.

FOLLICLE STIMULATING HORMONE. In the female, follicle stimulating hormone initiates growth of *follicles* (fluid chambers) in the ovaries. In each of these a single ovum develops in preparation for fertilization. This hormone also causes the ovaries to secrete *estrogens*, one of the female sex hormones. In the male, follicle stimulating hormone stimulates growth of the germinal epithelium in the testes, thus promoting the development of sperm that can then fertilize the female ovum.

LUTEINIZING HORMONE. In the female, luteinizing hormone causes the follicle to rupture, allowing the ovum to pass into the abdominal cavity and then through a fallopian tube, within which time fertilization may take place. In the male, luteinizing hormone causes the testes to secrete the male sex hormone *testosterone*.

LUTEOTROPIC HORMONE. In the female, luteotropic hormone causes the cells of the ruptured follicles to secrete *progesterone* as well as additional estrogens, which are necessary to prepare the uterine surface for implantation of a fertilized ovum. The function of luteotropic hormone in the male is yet unknown, though it probably has some important effect on either the production of male sex hormone or the production of sperm.

Regulation of adenohypophyseal secretion—the hypothalamic-hypophyseal portal system. The adenohypophysis is a highly vascular organ which receives blood supply from two sources: (1) the usual arteriolar source and (2) the so-called *hypothalamic-hypophyseal portal system* which is illustrated in Figure 318. The hypothalamus, a part of the brain lying immediately beneath the thalamus, controls many of the automatic functions of the body. After the blood passes through the capillaries of the hypothalamus, particularly through the lower part called the *median eminence*, it leaves by way of small *hypothalamic-hypophyseal portal veins* which course down the anterior surface of the hypophyseal stalk to the adenohypophysis. At this point the veins dip into the gland where the blood flows through large numbers of *venous sinuses* that bathe the adenohypophyseal cells.

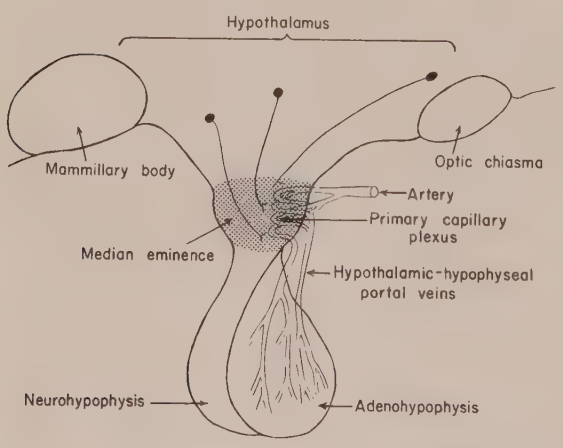


Figure 318. The hypothalamic-hypophyseal portal system.

The hypothalamus secretes a *neurosecretory substance* for each of the major hormones formed in the adenohypophysis. It is these neurosecretory substances passing to the adenohypophysis by way of the portal system that control secretion by the adenohypophyseal cells. For instance, the neurosecretory substance *thyrotropin releasing factor* causes the adenohypophysis to secrete large quantities of thyrotropin. Likewise, the neurosecretory substance *corticotropin releasing factor* causes the adenohypophysis to secrete large quantities of corticotropin. Similar neurosecretory substances control adenohypophyseal secretion of growth hormone and of each of the three gonadotropic hormones. Several of these neurosecretory substances are discussed in more detail in other parts of this text in relation to the functions of the thyroid gland, the adrenal cortex, and the sex glands.

The Neurohypophyseal Hormones

The neurohypophysis does not secrete any hormones at all but simply stores two hormones called *antidiuretic hormone* and *oxytocic hormone* which are secreted by cells in the anterior hypothalamus and conducted through nerve fibers into the neurohypophysis. Transmission of nerve impulses from the anterior hypothalamus to the neurohypophysis causes the stored hormones to be released.

Antidiuretic hormone and control of water reabsorption in the tubules of the kidney. The importance of antidiuretic hormone to the function of the kidney was discussed in Chapter 8. Briefly, antidiuretic hormone prevents the body fluids from becoming too concentrated in the following way: When the concentration of electrolytes in the fluids begins to rise, the increasing osmotic pressure of these fluids causes special neurons in the anterior hypothalamus, called *osmoreceptors*, to shrink, which excites them and sends impulses to the neurohypophysis. The impulses cause antidiuretic hormone to be released from the neurohypophysis. This in turn passes by way of the blood to the collecting tubules of the kidney and causes increased quantities of water to be

reabsorbed while allowing large quantities of electrolytes to pass on into the urine. Thus, the water of the body fluids is conserved, while the electrolytes gradually decrease. In this way, overconcentration of the extracellular fluids is corrected.

PRESSOR FUNCTION OF ANTIDIURETIC HORMONE. Antidiuretic hormone is also frequently called *vasopressin* because injection of very large quantities of purified hormone causes the arterial pressure to rise. However, only in rare instances is sufficient antidiuretic hormone *secreted* in the body to have significant effect on arterial pressure. Therefore, this pressor effect of antidiuretic hormone is more a pharmacological than a physiological effect.

Oxytocin. An "oxytocic" agent is a substance that causes the uterus to contract. This is one of the primary effects of *oxytocin*, also known as oxytocic hormone. This hormone is secreted in moderate quantities during the latter part of pregnancy and in especially large quantities at the time that the baby is to be born. It probably aids in the expulsion of the baby from the uterus. It is well known that a mother whose oxytocin secreting mechanism has been destroyed has considerable difficulty in delivering her baby. This will be discussed further in relation to childbirth in Chapter 37.

EFFECT OF OXYTOCIN ON MILK EJECTION. Oxytocin also has an important function in helping the mother to nurse the newborn infant as follows: Milk formed by the *glandular cells* of the breasts is secreted into the *alveoli* of the breasts where it is stored until the baby begins to nurse. For approximately the first minute after nursing begins the baby receives essentially no milk, but the suckling stimulus excites the mother's nipple and transmits impulses upward through the spinal cord, into the brain stem, and finally to the hypothalamus where they cause secretion of oxytocin. This hormone in turn flows by way of the blood to the breasts where it contracts many small *myoepithelial cells* surrounding the alveoli, thereby squeezing the milk into the ducts so that the baby can remove it by suckling.

THYROXINE

Formation of thyroxine by the thyroid gland. When stimulated by thyrotropin from the adenohypophysis, the thyroid gland, which is located in the neck beneath the larynx, forms *thyroxine* from *iodine* and *tyrosine*, one of the amino acids, by the following reaction:

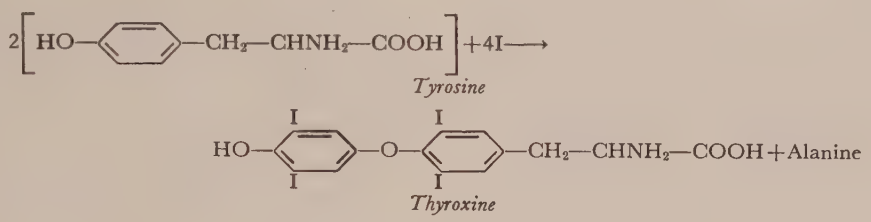


Figure 319 illustrates the basic structure of the thyroid gland, showing the follicular epithelium that secretes thyroxine and a colloid substance filling the follicles. The colloid contains large amounts of *thyroglobulin*, which is composed of thyroxine combined with a protein. Thyroxine is gradually released from the thyroglobulin into the blood as it is needed by the body. As much as several months' supply of thyroxine can be stored in the follicles, which means that one can live for many weeks on a diet deficient in iodine.

When thyroxine enters the blood it immediately combines with the plasma proteins. Gradually, small quantities of the thyroxine dissociate from these proteins and diffuse into the interstitial fluids and tissue cells. The transfer of

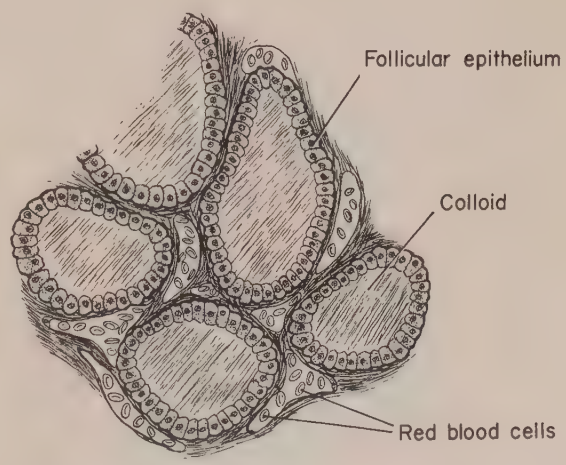


Figure 319. Cross-section of a portion of the thyroid gland, showing colloid material stored in the follicles.

thyroxine from the thyroid gland to tissue cells takes an average of three or more days, which assures a steady though slow flow of the hormone to the tissues. Because of this slowness, some of the effects of thyroxine are still apparent as long as six to eight weeks after its release from the thyroid gland.

Mechanism of Action of Thyroxine

Effect of thyroxine on cellular metabolism and on the cellular enzymes. Thyroxine increases the rate of metabolism of all cells. Though the precise means by which this is accomplished are not known, studies have shown that tissues exposed to thyroxine develop greatly enhanced quantities of most of their enzymes. To date, at least 13 different cellular enzymes have been shown to be greatly increased under the influence of thyroxine, and because enzymes are the regulators of chemical reactions in the cells it is quite easy to understand how this could increase the metabolic rates of the cells. Because thyroxine affects all tissues of the body, it regulates the overall rate of activity of all functions of the body.

Effects of thyroxine on some specific functions of the body. EFFECT ON TOTAL BODY METABOLISM. Total lack of thyroxine production by the thyroid gland decreases the metabolic rate to about one-half normal. On the other hand, secretion of very large quantities of thyroxine can increase the rate of metabolism to as much as 2 times normal. Therefore, in all, the thyroid gland can change the rate of metabolism as much as four-fold. Measurements of the basal metabolic rate can be used to estimate the degree of activity of the thyroid gland.

Thyroxine causes the body to burn its available carbohydrates very rapidly, and then to make deep inroads on the stores of fats. A person with excess production of thyroxine usually loses weight, sometimes very rapidly. On the other hand, a person with less than normal production of thyroxine often develops rather extreme obesity.

EFFECT ON THE CARDIOVASCULAR SYSTEM. Thyroxine affects the cardiovascular system in two ways. First, because the metabolic rate rises, all the tissues of the body require increased quantities of nutrients, and cardiovascular reactions, particularly vasodilatation in all the tissues, cause the heart to pump greater quantities of blood than usual. Second, thyroxine has a direct effect on the heart, increasing its metabolism as well as its rate of beat and forcefulness of contraction. Consequently the systolic arterial pressure rises above normal, and blood flow through the tissues increases still more.

EFFECT ON THE NERVOUS SYSTEM. Thyroxine greatly increases the reactivity of the nervous system. The reflexes become very excitable with excess production of thyroxine, but very sluggish with diminished production. Thyroxine increases a person's degree of wakefulness, while lack of thyroxine sometimes makes him sleep as much as 12 to 15 hours per day.

A special effect of thyroxine on the nervous system is a continuous tremor of the muscles. The tremor is usually very fine but rapid, having a frequency of 10 to 20 times per second, which is considerably faster than the tremor in basal ganglia or cerebellar disease.

EFFECT ON THE GASTROINTESTINAL TRACT. Thyroxine increases the motility of the gastrointestinal tract and promotes copious flow of digestive juices. If these activities are enhanced sufficiently, diarrhea may develop. On the other hand, lack of thyroxine causes the opposite effects —sluggish motility and greatly diminished secretion—resulting in constipation. Excess production of thyroxine also causes a voracious appetite because of the rapid rate of metabolism. One eats a tremendous amount of food, digests it rapidly, absorbs large quantities of nutrients, but metabolizes these as rapidly as they become available.

Regulation of Thyroxine Production

Earlier in the chapter it was pointed out that thyroxine production is regulated almost entirely by *thyrotropin* from the adenohypophysis. In turn, the rate of secretion of thyrotropin is regulated by *thyrotropin releasing factor* secreted by the hypothalamus. Therefore, to describe the regulation of thyroxine production, we need only to discuss the factors that regulate secretion of thyrotropin releasing factor.

The primary stimulus controlling the rate of secretion of thyrotropin releasing factor is the level of metabolism in the body. If the metabolism falls to a low value, the rate of secretion of thyrotropin releasing factor increases automatically, in turn increasing the secretion of thyrotropin and consequently of thyroxine. The thyroxine then increases the metabolic rate of the body back toward normal. On the other hand, if the body's rate of metabolism rises above normal, the hypothalamus decreases its secretion of thyrotropin releasing factor, and an opposite sequence of events reduces the secretion of thyroxine, thereby reducing the metabolic rate back toward normal. Thus, as demonstrated in

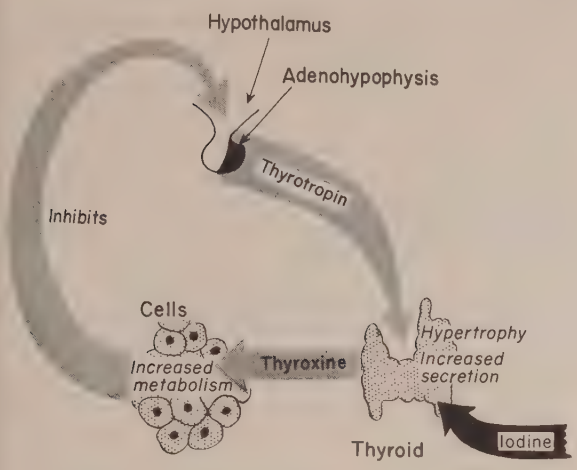


Figure 320. Interrelationships between the adenohypophysis and the thyroid gland for regulating the metabolic rate.

Figure 320, the thyrotropin-thyroxine mechanism acts as a feedback system to regulate body metabolism at a normal mean value.

Effect of cold weather on thyroxine production. When a person or an animal is exposed to very severe cold, the hypothalamus secretes increased quantities of thyrotropin releasing factor. Over a period of three to four weeks, the thyroid gland gradually enlarges, and the rate of thyroxine secretion increases. The basal metabolic rate can be increased by as much as 15 to 25 per cent by this mechanism, which helps to warm the body and partially compensates for the cooling effect of the weather.

Abnormalities of Thyroid Secretion

Hyperthyroidism. Failure of the adenohypophysis-thyroid regulatory system to function properly frequently leads to greatly increased production of thyroxine, sometimes to as much as 25 times normal. Almost always the increased production is caused by an increase in thyrotropin and not by an abnormality of the thyroid gland. But in a few persons a small *adenoma*, a tumor, develops in the thyroid gland and secretes thyroxine independently of control by the adenohypophysis. In either event the excess production of thyroxine is called *hyperthyroidism*.

In hyperthyroidism the basal metabolic rate rises very high—sometimes to as much as twice normal, the person loses weight, he develops diarrhea, he becomes highly nervous and tremulous, his heart rate is greatly increased, and the heart often beats so hard that he feels it palpitating in his chest. The state of hyperthyroidism is often so severe and so prolonged that it actually "burns out" the tissues, leading to degenerative processes in many parts of the body. One of the most common areas of degeneration is the muscle of the heart itself.

The usual methods for treating hyperthyroidism are (1) administration of a drug that suppresses thyroid function or destroys the thyroid gland, or (2) surgical removal of a major portion of the gland. Administration of *propylthiouracil* blocks the chemical reaction of tyrosine and iodine to form thyroxine, and hyperthyroidism can often be controlled with this drug. One of the more recent means for treating hyperthyroidism is the administration of radioactive iodine. Approximately two thirds of the ingested iodine is absorbed from the blood into the thyroid gland, so that a resulting intense radiation inside the gland destroys the overly active tissue, and normal thyroid function is reestablished.

EXOPHTHALMOS. Most patients with severe hyperthyroidism also develop protruding eyeballs, a condition called *exophthalmos* (Fig. 321). It is believed that thyroxine has nothing to do with the protrusion of the eyeballs but instead that this is caused by *exophthalmos producing substance* secreted in large quantities by the adenohypophysis along with the excessive secretion of thyrotropin that occurs in hyperthyroidism. The cause of this association between exophthalmos producing substance and thyrotropin is not understood, though exophthalmos is indeed a very unfortunate and common side effect of hyperthyroidism. Exophthalmos producing substance causes edema, excessive growth, and degeneration of the tissues behind the eyeballs. Because exophthalmos results at least partially from increased quantity of tissue behind the eyes,



Figure 321. A hyperthyroid person with exophthalmos.

elimination of the hyperthyroidism will not eliminate all the exophthalmos.

Hypothyroidism. Diminished production of thyroxine is called *hypothyroidism.* A person can live for many years with complete lack of thyroxine production, but the rate of metabolism in all his tissues is decreased to about one-half normal. He is extremely lethargic, sleeping sometimes as much as 12 to 15 hours a day. He usually is constipated, his mental reactions are sluggish, and he often becomes fat. In addition to increased deposition of fat throughout his body, a gelatinous mixture of protein and extracellular fluid deposits in the spaces between the cells, giving an edematous appearance. For this reason the condition is often called *myxedema.*

Goiter. In hyperthyroidism the overactive thyroid gland usually enlarges two- to three-fold, and is then called a *goiter.* In hypothyroidism the gland frequently enlarges also, and here again the enlarged gland is called a goiter. Therefore, the term goiter is not synonymous with either hyper- or hypothyroidism.

Hypothyroidism is usually caused by some abnormality of the thyroid gland that makes it impossible for the gland, even when stimulated with thyrotropin, to secrete enough thyroxine. Yet the poorly secreting gland enlarges more and more in a futile attempt to produce adequate quantities of thyroxine, and large amounts of colloid substance containing almost no thyroxine are secreted into the follicles. For this reason this type of enlarged thyroid gland is called a *colloid goiter.*

Sometimes a colloid goiter becomes 15 times as large as the normal thyroid gland, weighing 500 or more grams and occupying a space in the neck or upper chest as large as one-half liter. Obviously, a gland this large can obstruct breathing and swallowing.

ENDEMIC GOITER. Persons residing in areas of the world where the food contains very little iodine cannot produce an adequate quantity of thyroxine. As a result, their metabolic rates decrease below normal, and this enhances the output of thyrotropin which in turn stimulates the thyroid gland in an attempt to produce increased quantities of thyroxine. Unfortunately, even this stimulus cannot enhance the output of thyroxine when iodine is lacking. But the adenohypophysis continues to produce large amounts of thyrotropin, so that the thyroid continues to enlarge, becoming progressively filled with colloid that contains almost no thyroxine. The enlarged gland is called an *endemic goiter* because everyone in the area develops an enlarged gland. Endemic goiter was formerly widely prevalent in those regions of the world, such as the Great Lakes region of the United States and the Swiss Alps, where iodine is not present in the soil. More recently, however, a small amount of iodine has been added to most commercial table salts, so that now an inadequate intake of iodine is very rare.

REFERENCES

Astwood, E. B.: *Progress in Clinical Endocrinology.* Vol. II. New York, Grune & Stratton, Inc., 1958.

Fortier, C.: Adenohypophysis and adrenal cortex. *Ann. Rev. Physiol.* 24:223, 1962.

Hoch, F. L.: Biochemical actions of thyroid hormones. *Physiol. Rev.* 42:605, 1962.

Ingbar, S. H., and Galton, V. A.: Thyroid. *Ann. Rev. Physiol.* 25:361, 1963.

Kleeman, C. R., and Cutler, R. E.: The neurohypophysis. *Ann. Rev. Physiol.* 25:385, 1963.

Paschkis, K. E., Rakoff, A. E., and Cantarow, A.: *Clinical Endocrinology.* 2nd Ed. New York, Paul B. Hoeber, Inc., 1958.

Smith, R. W., Jr., Gaebler, O. H., and Long, C. N. H.: *The Hypophyseal Growth Hormone, Nature and Actions.* New York, McGraw-Hill Book Company, Inc., 1955.

Soderberg, U.: Temporal characteristics of thyroid activity. *Physiol. Rev.* 39:777, 1959.

Van der Kloot, W. G.: Comparative endocrinology of the invertebrates. *Ann. Rev. Physiol.* 24:491, 1962.

Werner, S. C.: *The Thyroid.* New York, Paul B. Hoeber, Inc., 1956.

Williams, R. R.: *Textbook of Endocrinology.* 3rd Ed. Philadelphia, W. B. Saunders Company, 1962.

ADRENOCORTICAL HORMONES AND INSULIN

34

ADRENOCORTICAL HORMONES

The adrenal glands, shown in cross-section in Figure 322, are elongated structures located immediately above each of the two kidneys. These glands produce hormones of two entirely different types. The *medulla* of each gland is part of the sympathetic nervous system, and it secretes *epinephrine* and *norepinephrine* in response to sympathetic stimuli; this was described in Chapter 26. The outer portion of the adrenal gland is the *cortex,* which secretes *adrenocortical hormones.* These hormones are all chemically similar, but they can be divided into three different categories on the basis of their functions: *mineralocorticoids, glucocorticoids,* and *androgens.* The mineralocorticoids control the electrolyte balances of sodium, chloride, and potassium; the glucocorticoids affect metabolism of protein, fat, and glucose; and the androgens cause masculinizing effects.

Mineralocorticoids

The function of mineralocorticoids is to regulate electrolytes, particularly sodium, in the extracellular fluids.

The adrenal cortex secretes at least three different hormones that can be classified as mineralocorticoids: aldosterone, corticosterone, and minute quantities of deoxycorticosterone. However,

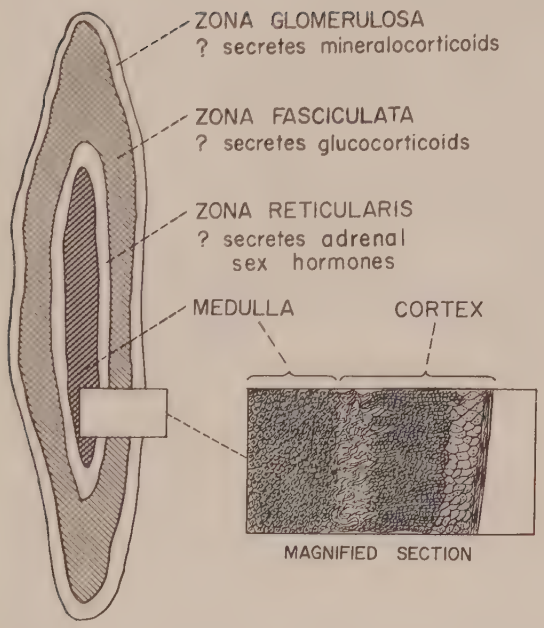


Figure 322. Cross-section of the adrenal gland, showing the three zones of the adrenal cortex that are believed to secrete the three types of adrenocortical hormones.

aldosterone accounts for at least 95 per cent of the total mineralocorticoid activity.

The function of mineralocorticoids in relation to electrolyte absorption from the kidney tubules has already been presented in Chapter 8; their overall function can be summarized briefly as follows:

Effect on sodium. A direct effect of aldosterone is an increase in the rate of sodium reabsorption by the renal tubular epithelium. When large quantities of aldosterone are secreted, essentially all of the sodium entering the glomerular filtrate is reabsorbed into the blood, and almost no sodium passes into the urine. On the other hand, when minute amounts of aldosterone are secreted, much less sodium is reabsorbed, and as much as 20 to 30 grams of sodium may appear in the urine each day. Therefore, it is evident that aldosterone is essential to prevent rapid depletion of sodium in the body. Thus, aldosterone secretion represents a method for controlling the total quantity of sodium in the extracellular fluids.

Effect on potassium. The enhanced sodium reabsorption from the tubules is accompanied by decreased reabsorption of potassium. This probably results from the following mechanism: When excessive amounts of sodium are reabsorbed under the influence of aldosterone, the positive charges of the sodium ions create electropositivity in the extracellular fluids surrounding the tubules. This electropositivity in turn opposes reabsorption of the positively charged potassium ions. As a result, an increased amount of potassium flows on through the tubules into the urine.

Effect on chloride. Since sodium is by far the major alkaline ion of the extracellular fluid, enhanced reabsorption of sodium causes the extracellular fluid to become slightly alkaline. The alkaline reaction in turn increases the reabsorption of chloride by the kidney tubules, as explained in Chapter 8. Thus, another secondary effect of mineralocorticoid action is a decrease in the amount of chloride lost into the urine, and an increase in the concentration of chloride in the extracellular fluid. The net result is an increased concentration of sodium chloride (salt) in the extracellular fluid and a slightly decreased concentration of potassium.

Effect on water reabsorption in the kidneys. The increased concentration of sodium chloride in the extracellular fluids has two effects on the absorption of water from the kidney tubules. First, absorption of sodium and chloride from the tubules diminishes the concentration of these substances in the tubular fluid and diminishes the tubular crystalloidal osmotic pressure. This allows increased absorption of water by osmosis into the peritubular fluids. Second, the increased concentration of sodium and chloride in the extracellular fluids increases the osmotic pressure of these fluids, which stimulates the osmoreceptor antidiuretic mechanism for enhancing the reabsorption of water by the tubules. This was explained in the previous chapter. Thus, by two means the quantity of water absorbed from the tubules is increased, and the total volume of extracellular fluid becomes greatly enhanced, sometimes as much as 100 per cent.

Effect on fluid volumes and circulatory dynamics. The increased extracellular fluid enters both the interstitial spaces and the blood, *increasing the interstitial fluid volume and the blood volume.* The increased blood volume forces a greater than normal quantity of blood toward the heart, increasing the cardiac output; for reasons that are not completely understood, the vascular resistance in the peripheral circulatory system increases too. This combination causes an *elevated arterial pressure.* Therefore, because of the effect on sodium absorption in the kidney tubules, aldosterone can indirectly affect blood volume, cardiac output, and arterial pressure.

Regulation of aldosterone secretion. Any one of three different factors can cause the rate of aldosterone secretion to increase: low extracellular fluid sodium concentration, high extracellular fluid potassium concentration, and low extracellular fluid volume.

The mechanism by which these stimuli cause the adrenal cortex to secrete increased aldosterone is not completely understood. One theory is the following: Any one of these three stimuli supposedly can excite nervous centers in the diencephalic region of the brain, probably in nervous centers near the pineal gland or perhaps even the pineal gland itself, causing the production of a substance called *glomerulotropic hormone.* This hormone then flows with the blood to the adrenal cortex where it stimulates the outer zone, the *zona glomerulosa,* to secrete large quantities of aldosterone. A second theory is that one or more of the above factors might cause decreased arterial pressure which in turn causes the kidneys to secrete a humoral substance called *aldosterone stimulating substance.* This substance is believed to be renin, and it supposedly excites the adrenal cortex to secrete aldosterone. Both of these theories have many objections so that it is possible that neither is correct.

The aldosterone in turn causes increased reabsorption of sodium, increased loss of potassium into the urine, and increased extracellular fluid. Therefore, here again is a feedback mechanism, regulating extracellular fluid volume and regulating the ratio of sodium to potassium in the extracellular fluids.

Glucocorticoids

The precise functions of the glucocorticoids are not nearly as well understood as those of the mineralocorticoids. However, metabolic systems become greatly deranged when glucocorticoids are absent from the body, and one becomes unable to resist almost any traumatic or disease condition that tends to destroy tissues. Therefore, the most important function of glucocorticoids is to enhance resistance to physical "stress," though the means by which this is effected are yet very vague. Several different adrenocortical hormones exhibit glucocorticoid activity, but by far the most abundant one is *hydrocortisone.* Other much less important glucocorticoids are cortisone and corticosterone. The latter is a mineralocorticoid as well as a glucocorticoid.

Effect of hydrocortisone on protein metabolism. Even though the name glucocorticoid implies that this type of adrenal hormone affects mainly glucose metabolism, it is becoming apparent that its most important action is on protein metabolism. Administration of hydrocortisone causes an increase in the amount of amino acids in the extracellular fluid and an increased rate of utilization of these acids for various purposes. The mechanism by which hydrocortisone increases the extracellular amino acids is not completely known, but experiments have shown that these hormones increase the rate of transfer of amino acids out of the cells into the blood. This makes the amino acids available for use wherever they are needed to provide energy, to repair damaged tissues, or even to synthesize necessary structures in new cells. This is a possible explanation for the ability of hydrocortisone and other glucocorticoids to enhance resistance to factors that destroy tissues.

In summary, even though the precise effects of glucocorticoids on protein metabolism are yet unknown, it seems that by increasing the transport of amino acids through cellular membranes these hormones increase the flux of amino acids throughout the body, allowing them to pass from one part of the body to another with increased ease, and especially allowing them to flow in great quantity to those areas where they are momentarily in special demand.

Effect of hydrocortisone on fat metabolism. Hydrocortisone mobilizes fat from the fat depots in much the same manner as it mobilizes amino acids from the cells. The net result is a decrease in the amount of fats in the storage areas and increased use of fats for energy and other purposes. It is believed, for example, that the mobilization of fats from the storage areas during periods of starvation is caused by an increase in the production of hydrocortisone.

During periods of rapid fat mobilization, the liver splits the fat into keto acids. If these are used immediately by the cells for energy, they cause no significant physiologic effect, but if they are not used immediately, their concentration in the extracellular fluid can become great enough to cause acidosis. This is often one of the untoward effects of excessive hydrocortisone secretion.

Effect of hydrocortisone on glucose metabolism. The earliest discovered effect of glucocorticoids was their ability to increase the concentration of glucose in the blood. Subsequent study indicates that this ability is actually an effect on protein and fat metabolism, for the increased glucose concentration is caused by increased conversion of proteins and, to a less extent, fats into glucose, a process called *gluconeogenesis.* It is believed that when protein and fat are mobilized from their storage areas, their increased concentrations in the body fluids provide increased quantities of material for the liver to convert into glucose. Regardless of the exact mechanism by which gluconeogenesis occurs, this function of hydrocortisone is very important. It keeps the blood glucose concentration high even

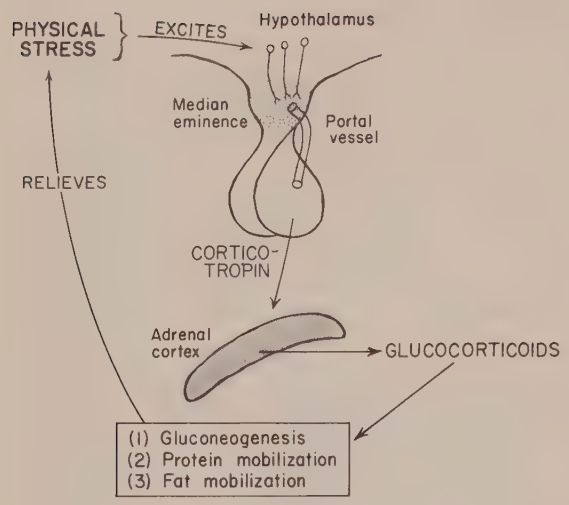


Figure 323. Mechanism by which physical stress induces adrenal secretion of hydrocortisone and other glucocorticoids.

during starvation, and thereby provides adequate nutrition for the neurons that can use only glucose for energy.

Regulation of hydrocortisone secretion. Figure 323 summarizes the regulation of adrenal secretion of hydrocortisone and other glucocorticoids. The mechanism can be explained as follows: The primary stimulus that initiates glucocorticoid secretion is called "stress," which includes almost any physical damage to the body. For instance, a painful contusion of some part of the body, a broken bone, severe damage to large areas of cells by some diseased condition, or any other destruction of parts of the body sets off the sequence of events that leads to hydrocortisone secretion. The stress probably causes these reactions by initiating nerve impulses that are transmitted from the periphery into the hypothalamus. The hypothalamus then secretes a substance called *corticotropin releasing factor* which passes by way of the hypothalamic-hypophyseal portal system into the adenohypophysis. Here, this factor causes the cells of the adenohypophysis to secrete *corticotropin* which flows in the blood to the adrenal cortex where it elicits hydrocortisone secretion and to a much less extent secretion of other glucocorticoids. The glucocorticoids then mobilize protein and fat from all over the body and cause gluconeogenesis. The increased availability of amino acids, fats, and glucose helps in the repair of the physical damage, thus relieving the initial stimulus that had set off the sequence of events leading to hydrocortisone secretion. Thus, this is also a feedback mechanism, which allows rapid mobilization of the necessary metabolic building blocks required for repair of tissues whenever physical damage occurs.

Androgens

Androgens are hormones that cause male sex characteristics, as we shall see in Chapter 36. Though the male testes are the primary source of these hormones, the adrenal cortex also secretes minute quantities, quantities so small that in the normal person they have no significant effect. However, an androgen-producing tumor of the adrenocortical cells develops occasionally which secretes very large quantities of male sex hormones that cause serious masculinizing effects in the body. This is explained below.

Abnormalities of Adrenocortical Secretion

Hyposecretion of adrenocortical hormones—Addison's disease. The adrenal cortices are occasionally destroyed by disease, or sometimes they simply atrophy. Tuberculosis frequently destroys the adrenal cortices. Occasionally, excessive overstimulation of the adrenal gland by stress causes it to become, first, greatly enlarged, then hemorrhagic, and finally replaced almost completely by fibrous tissue. In each instance hyposecretion of adrenocortical hormone occurs, or complete lack of secretion, leading to a serious condition called *Addison's disease.*

Complete failure of the adrenal cortices usually causes death within three to five days unless appropriately treated. This early death is caused by the lack of aldosterone, because the kidney's sodium reabsorptive mechanism is almost completely dependent on this hormone. Without sodium reabsorption the level of sodium in the extracellular fluid and the volume of extracellular fluid decrease so greatly within only a few days that death ensues. This effect can be offset to some extent by having the afflicted person eat extra quantities of salt, and it can be overcome entirely by administering one of the mineralocorticoids; the one most usually used because of its easy synthesis is the compound *deoxycorticosterone.*

Even if the life of the person with Addison's disease is saved by administering deoxycorticosterone, he still remains unable to resist stress, and even a slight respiratory infection may prove lethal. He also has little energy. Usually, therefore, to provide completely adequate function of the body, treatment with hydrocortisone or some other glucocorticoid is necessary in addition to deoxycorticosterone.

Hypersecretion of adrenocortical hormones. Hypersecretion of the adrenal cortex can result either directly from a tumor in one part of an adrenal gland, or indirectly from increased pro-

duction of corticotropin by the adenohypophysis, in turn stimulating the adrenal cortices.

The effects of hyperadrenalism depend upon which part of the adrenal gland is secreting the excessive quantities of hormone. If it is the inner zone of the adrenal cortex, the *zona reticularis,* then excessive quantities of androgens are secreted, and the person, even a child or a female, develops masculine characteristics such as growth of hair on the face, deepening of the voice, sometimes baldness, changes in portions of the female sexual organs to resemble the sexual organs of the male, atrophy of the female breasts, and considerably enhanced muscular development.

If the middle zone of the cortex, the *zona fasciculata,* secretes excess hormones, there are symptoms of excess hydrocortisone secretion. These include excessive mobilization of proteins and fats from their storage areas. The mobilization of proteins causes weakness of the muscles and sometimes weakness of other protein structures, such as the fibers that hold the tissues together beneath the skin. This allows laxity of the skin and predisposes to tears in the subcutaneous tissues that can be observed as long, linear, purplish *striae.* The excess mobilization of proteins and fats also increases gluconeogenesis, and raises the blood glucose concentration, sometimes enough to cause a very high blood sugar level, a condition known as *pituitary diabetes.*

If the increased secretion occurs in the outer layer of the adrenal cortex, the *zona glomerulosa,* the effects are those of excessive aldosterone secretion. The concentration of sodium chloride in the extracellular fluid increases, the concentration of potassium decreases, the blood volume increases, cardiac output increases, and the arterial pressure increases.

INSULIN

Secretion of insulin by the pancreas. The pancreas, a section of which is shown in Figure 324, is composed of two different types of tissue. One type is the *acini,* which secrete digestive juices into the intestines. The other type is the *islets of Langerhans,* which secrete *insulin* and one or more less important hormones directly into the body fluids; this secretion into the body fluids is an *endocrine* function of the pancreas, while the secretion of digestive juices is an *exocrine* function.

The islets of Langerhans are composed of two different types of cells, the *alpha* and *beta cells.* The beta cells secrete insulin, while the alpha cells are believed to secrete a hormone called *glucagon,* which is discussed later in the chapter.

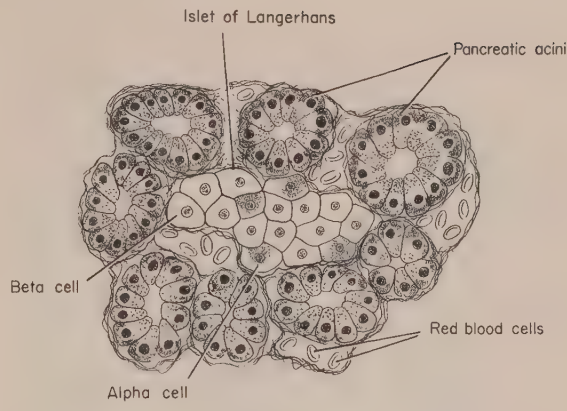


Figure 324. Microscopic structure of the pancreas.

Mechanism of Insulin Action

The basic mechanism of insulin action is to increase the rate of glucose transfer through the cellular membrane. The pores of the membrane are much too small for the glucose molecule to enter the cell by the process of diffusion. Instead, glucose must be carried through the membrane by an active transfer process such as that shown in Figure 325. In the absence of insulin only a small amount of glucose can be transported into the cells, but in the presence of normal amounts of insulin the transfer is accelerated as much as 3- to 5-fold, and in the presence of large amounts of insulin the rapidity of glucose transfer is increased as much as 15- to 25-fold. Therefore, insulin controls the rate of glucose metabolism in the body by controlling the entry of glucose into the cells.

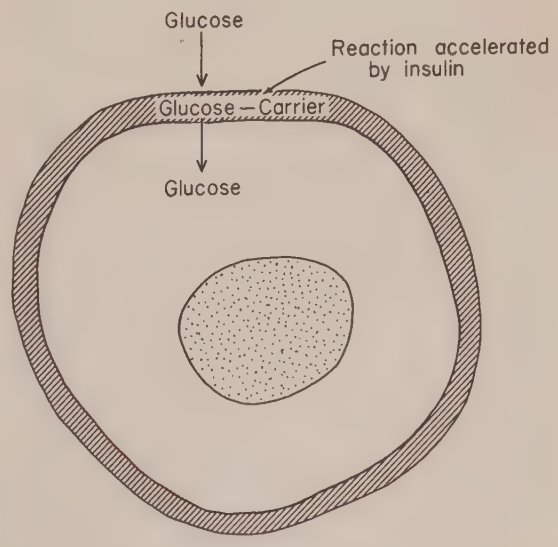


Figure 325. Basic mechanism of insulin action.

Among the recent research findings that have proved this concept—that the major action of insulin is to control the transport of glucose through the cellular membrane—are the following: (1) Cells removed from the body will utilize only minimal amounts of glucose from the surrounding fluids in the absence of insulin. But if these cells are crushed so that their membranes are destroyed, the enzyme systems of the cells begin very rapid utilization of the glucose. This indicates that the enzymes of the cells can utilize glucose very rapidly even in the absence of insulin if the glucose can reach the enzymes. (2) Insulin has been shown to increase the membrane transport of glucose and other monosaccharides such as galactose which has a chemical structure very similar to that of glucose. Galactose is far more easily studied than glucose because it is not metabolized once it enters the cells. The fact that insulin increases its entry into the cell is indication that insulin also promotes the transport of glucose through the membrane. (3) Studies have indicated that insulin is concentrated in the cellular membrane. This, too, indicates that its action is probably at the membrane or certainly near the membrane.

Effect of insulin on blood glucose. Because insulin accelerates the rate of glucose transfer from the extracellular fluids to the interior of the cells, the concentration of glucose in the blood and extracellular fluids becomes diminished. Conversely, lack of insulin secretion causes glucose to dam up in the blood instead of entering the cells. Complete lack of insulin usually produces a rise in blood glucose concentration from a normal value of 90 mg. per 100 ml. up to about 350 mg. per 100 ml. On the other hand, a great excess of insulin can decrease the blood glucose to about 25 mg. per 100 ml., which is about one-fourth normal.

Effect of insulin and insulin lack on fat metabolism. Insulin probably has no direct effect on fat metabolism, but it does have exceedingly important secondary effects. When glucose cannot be utilized for energy because of the absence of insulin, the metabolic mechanisms of the body must then utilize fat for energy—fat obtained from either the diet or the fat depots. The enhanced use of fat increases the blood concentrations of most of the intermediary products of fat metabolism, including especially the keto acids, cholesterol, and other blood lipids. The keto acids often cause acidosis, and the excess cholesterol and other lipids lead to atherosclerosis, which is discussed later in this chapter.

Excess insulin induces exactly opposite effects on fat metabolism. The insulin causes all the available glucose to enter the cells very rapidly, and conversion of the glucose into fats is greatly accelerated, which can be explained as follows: The amount of energy used by the cells is regulated by the rate of adenosine triphosphate breakdown and not by the amount of energy-giving substances available to the cell. Therefore, if an excess quantity of glucose enters the cell, only that portion needed immediately for energy is oxidized, and the remainder is converted into fat.

Effect of insulin on protein metabolism. Protein metabolism, like fat metabolism, probably is not affected directly by insulin, but indirectly it too is altered greatly by the action of insulin. When glucose cannot be utilized for energy, its *protein-sparing* effect is lost, and large quantities of protein are often utilized along with fats for energy in place of carbohydrates. As a consequence, fewer amino acids become available for synthesis of cellular and intercellular structures. For this reason, lack of insulin retards the growth of an animal and also retards the repair of damaged tissues. In fact, one of the serious consequences of insulin lack (*diabetes*) is diminished resistance to infections, trauma, and other types of physical stress.

Regulation of Insulin Secretion

When the blood glucose level rises, the pancreas begins secreting insulin within a few minutes. This is probably caused by a direct effect of glucose on the pancreas. That is, a high concentration of glucose automatically stimulates the beta cells to produce increased quantities of insulin. The insulin causes the excess glucose to be transported into the cells where it can be used for energy, stored as glycogen, or converted into fat.

Glucose tolerance test. A method often used for testing the ability of the pancreas to respond to glucose stimulation is the so-called glucose tolerance test, illustrated in Figure 326. In the normal person the blood glucose concentration at the beginning of the test is less than 100 mg. per cent. At this time 50 gm. of glucose is ingested, and during the ensuing half hour or more the blood glucose rises to a value of 130 to 160 mg. per cent. In less than one hour, however, large quantities of insulin will have been produced by the pancreas causing the blood glucose concentration to fall very rapidly, reaching normal once again about two hours after the initial ingestion of glucose. Then the concentration actually falls below normal, which is called the *hypoglycemic response,* because of continuing action of the

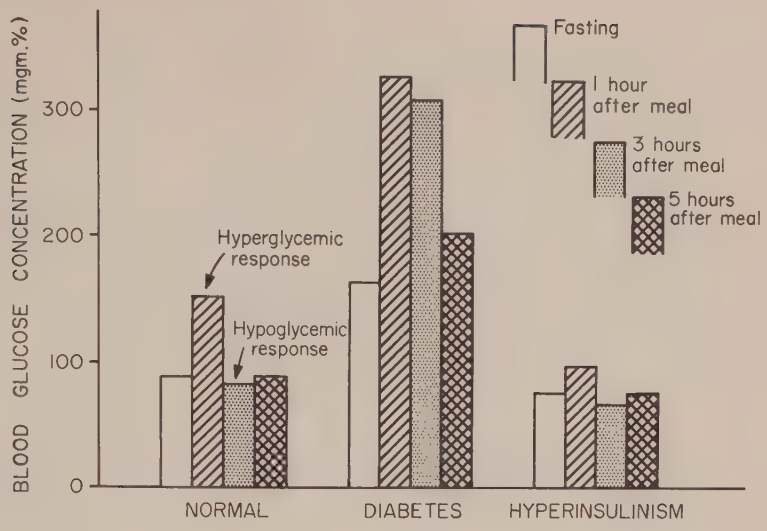


Figure 326. Effect on blood glucose concentration of administering 50 gm. of glucose to, first, a normal person, second, a diabetic person, and, third, a person with hyperinsulinism.

insulin secreted during the hyperglycemic phase of the test.

A diabetic person whose pancreas cannot secrete insulin often exhibits a hyperglycemic response in which the blood glucose concentration rises to as high as 200 or more mg. per 100 ml., and the blood glucose level may require as long as five or more hours to return to normal. Also, because no insulin has been secreted, the glucose level never shows the hypoglycemic response.

Diabetes

Diabetes is the disease that results from failure of the pancreas to secrete insulin. Usually it is caused by degeneration of the beta cells of the islets of Langerhans, though the cause of degeneration is usually not known. In lower animals certain chemical compounds, notably *alloxan,* can cause degeneration of these cells, and excess dietary carbohydrates in some instances can so overwork the secretory function of the beta cells that they actually "burn out." Therefore, it is possible that some persons develop diabetes secondary to exposure to abnormal chemicals, or secondary to eating excess quantities of carbohydrates. However, most diabetics are offspring of one or two parents who also had diabetes. Consequently, it is believed that the usual cause of diabetes is hereditary degeneration of the beta cells. This occurs in some persons at a very early age and in others at older ages.

Effects of diabetes. The primary abnormality in diabetes is failure to utilize adequate quantities of glucose for energy. This causes the blood glucose level to rise to as high as three times normal. Large quantities of glucose are lost into the urine because the kidney tubules cannot reabsorb all that enters them in the glomerular filtrate each minute. The glucose that is not reabsorbed by the tubules also creates a tremendous amount of crystalloidal osmotic pressure in the tubules and diminishes the reabsorption of water. As a result, the diabetic person loses large quantities of water as well as glucose into the urine. In extreme cases the excess output of urine causes extracellular dehydration, which can in itself be very harmful.

Failure of the diabetic person to utilize glucose for energy deprives him of a major portion of the energy in his food. He loses weight and becomes weakened because of excess burning of his fat and protein stores. As a result of the nutrient deficiency in diabetes, the diabetic person usually becomes very hungry, so that he often eats voraciously even though the carbohydrate portion of his food contributes little to his nutrition.

KETOSIS AND DIABETIC COMA. The extremely rapid metabolism of fats in diabetes sometimes increases the quantity of keto acids in the extracellular fluids to as high as 20 to 30 milliequivalents per liter, which is 25 to 50 times normal. On occasion this becomes sufficient to make the pH of the body fluids fall from its normal value of 7.4 to as low as 7.0, or rarely to as low as 6.9. This degree of acidosis is incompatible with life for more than a few hours. The person breathes extremely rapidly and deeply to blow off carbon dioxide, which helps to offset the metabolic acidosis, but despite this the acidosis often becomes severe enough to cause coma, and unless the person is treated he usually dies in less than 24 hours. Treatment requires immediate administration of extreme quantities of insulin. Sometimes glucose is administered along with the

insulin to help promote the shift from fat to glucose metabolism. Intravenous administration of alkaline solutions can be of great benefit for two reasons: First, alkaline solutions neutralize some of the acidosis, and, second, they correct the dehydration that often accompanies severe diabetic coma.

Treatment of diabetes with insulin. The diabetic person can usually be treated quite adequately by daily injections of insulin. Two types of insulin available for this purpose are *crystalline zinc insulin* and *protamine zinc insulin.* The duration of action of crystalline zinc insulin after injection is 4 to 6 hours, and that of protamine zinc insulin is 24 to 30 hours. Usually a person who has severe diabetes must take an injection of crystalline zinc insulin at meal times when his blood glucose concentration is likely to rise very high temporarily, and he must take an injection of protamine zinc insulin each morning to provide a steady rate of glucose inflow into his cells throughout the day.

Atherosclerosis in diabetes. Prolonged diabetes usually leads to early development of atherosclerosis, and this subsequently causes heart disease, kidney damage, cerebral vascular accidents, or other circulatory disorders. The probable reason for the development of atherosclerosis is that, even with the best possible treatment of diabetes, glucose metabolism can never be maintained at a sufficiently high level to prevent some excess fat metabolism, and cholesterol deposition in the walls of the blood vessels is always an unfortunate accompaniment of rapid fat metabolism. Because of this fact the person who develops diabetes in childhood usually has a shortened life, regardless of how well he is treated.

Hyperinsulinism

Hyperinsulinism occasionally develops because of either too much insulin injected or too much insulin secreted by a pancreatic islet tumor. In either case the low blood glucose concentration causes overexcitability of the brain at first and then coma. The neurons require a constant supply of glucose, because they cannot utilize fats or proteins for energy. Furthermore, the rate of glucose uptake by the neurons, unlike other cells, is dependent on the blood glucose concentration rather than on the amount of insulin available. Whenever excess insulin is available, too much of the blood glucose passes into all cells except the neurons, and the quantity of glucose in the extracellular fluid becomes very low. Then the neurons no longer receive the amounts of glucose needed to maintain their metabolism. This causes them first to become excessively excitable, and later depressed. In the excitement stage convulsions may occur, but in the depressed stage the person develops coma not unlike that of diabetes. Indeed, it is sometimes a problem to diagnose the cause of coma in a diabetic. It may result from too little insulin secretion, in which case it is diabetic coma, or from too much insulin, in which case the abnormality is hyperinsulinism.

Glucagon

It is believed that the alpha cells of the islets of Langerhans secrete a hormone called *glucagon.* The action of glucagon is to cause *glycogenolysis* (breakdown of glycogen to glucose) in the liver cells. This allows glucose to be dumped from the liver into the blood and extracellular fluids. The blood glucose concentration rises many milligrams per 100 ml. within a few minutes after injection of glucagon.

The value of glucagon in the body is completely unknown, because the method by which its secretion is regulated has not yet been determined. It is possible that when large quantities of glucose are suddenly needed by the tissue cells, the pancreas, for reasons yet unknown, may release glucagon. This then could establish a high concentration of glucose in the blood if the glycogen stores in the liver are sufficient.

REFERENCES

Duncan, G.: *Symposium on Diabetes.* New York, Grune & Stratton, Inc., 1957.

Flink, E. B.: Adenohypophysis and adrenal cortex. *Ann. Rev. Physiol.* 23:229, 1961.

Forsham, P.: *Symposium on Adrenal Steroids.* New York, Grune & Stratton, Inc., 1958.

Gardiner-Hill, H.: *Modern Trends in Endocrinology.* New York, Paul B. Hoeber, Inc., 1958.

Gardner, L. I.: *Adrenal Function in Infants and Children.* New York, Grune & Stratton, Inc., 1956.

Root, H. F., and White, P.: *Diabetes Mellitus.* New York, McGraw-Hill Book Company, Inc., 1956.

Stadie, W. C.: Current concepts of the action of insulin. *Physiol. Rev.* 34:52, 1954.

Vallance-Owen, J., and Wright, P. H.: Assay of insulin in blood. *Physiol. Rev.* 40:219, 1960.

Yates, F. E., and Urquhart, J.: Control of plasma concentrations of adrenocortical hormones. *Physiol. Rev.* 42:359, 1962.

Zimmermann, B.: *Endocrine Functions of the Pancreas.* Springfield, Ill., Charles C Thomas, 1952.

CALCIUM METABOLISM, BONE, PARATHYROID HORMONE, AND PHYSIOLOGY OF TEETH

35

CALCIUM METABOLISM

The adult human body contains about 1200 gm. of calcium, at least 99 per cent of which is deposited in the bones, but a very small and extremely important portion of which is in the blood plasma and interstitial fluid. Approximately one half of that in the plasma is ionized, and the remaining half is bound with the plasma proteins. It is the ionized calcium that diffuses into the interstitial fluid and enters into chemical reactions; it is this portion of the extracellular fluid calcium that is most important.

Functions of calcium ions. One of the principal functions of calcium ions occurs at the cellular membranes. In unicellular animals, calcium increases the thickness and strength of the membrane, and without calcium the membrane becomes very thin and friable (easily ruptured or crumbled). In the human being this basic effect of calcium on the membrane is not so obvious, but some of the secondary effects of its action on the cellular membrane are very important. For instance, lack of calcium causes the membranes of nerve fibers to become partially depolarized, and, therefore, to transmit repetitive and uncontrolled impulses. These may occur so rapidly that

they actually cause spasm of the skeletal muscles, a condition called *tetany*. On the other hand, a great increase in the concentration of calcium ions depresses the neurons in the central nervous system. This presumably occurs because the membranes will not depolarize with normal ease.

A second effect of decreased calcium ion concentration is weakness of the heart muscle. Decreased calcium causes the duration of cardiac systole to decrease, and the heart dilates excessively during diastole. Usually, however, the person dies of tetany before his heart function is greatly impaired by calcium deficiency. An excess of calcium promotes overcontraction of the heart, causing the muscle to contract much too forcefully during systole and not to relax satisfactorily during diastole. Fortunately, this effect of excess calcium can be demonstrated only in experiments, for it never occurs to a severe extent even in disease conditions.

These effects of high and low calcium concentrations on the heart can be explained by the basic mechanism of muscle contraction which was discussed in Chapter 21. When the cardiac impulse passes over the cardiac muscle, a small amount of calcium flows through the muscle membrane to the inside, and the calcium ion sets off the contractile process. When there are extremely small amounts of calcium in the extracellular fluids, the intensity of contraction is reduced, whereas excess calcium in the extracellular fluids causes overcontraction of the heart.

Another important function of calcium ions is the promotion of blood coagulation. It will be recalled from the discussion of blood coagulation in Chapter 11 that calcium enters into the chemical reaction between prothrombin, thromboplastin, and other factors to cause the formation of thrombin. The thrombin in turn catalyzes the formation of fibrin threads from fibrinogen, and thereby causes the blood to clot. Fortunately, the calcium ion concentration almost never falls low enough nor rises high enough to cause abnormalities of clotting.

Finally, an extremely important function of calcium ions is to react with phosphate ions to form bone salts, a function that is discussed in detail in subsequent parts of this chapter.

Reaction of Calcium Ions with Phosphate Ions

Calcium (Ca^{++}) and phosphate (HPO_4^{--}) ions react together to form *calcium phosphate* ($CaHPO_4$), a relatively insoluble compound. The product of the concentrations of calcium and phosphate ions in a solution can never be greater than a critical value called the *solubility product*, or they will precipitate in the form of calcium phosphate crystals. The greater the concentration of calcium in the solution, the less can be the concentration of phosphate, and the greater the concentration of phosphate, the less can be the concentration of calcium.

One of the places where the precipitation of calcium and phosphate is especially important is in the gastrointestinal tract, for the simultaneous presence of both of these substances in the food often leads to precipitation of large quantities of calcium phosphate, which cannot be absorbed from the gut with ease.

Another instance in which calcium and phosphate are inextricably related to each other is in the deposition and reabsorption of bone, because the salts of bone are mainly calcium phosphate compounds. Every time calcium is deposited phosphate is deposited also, and every time bone is reabsorbed both calcium and phosphate enter the body fluids at the same time.

Function of Vitamin D in Calcium Metabolism

Vitamin D greatly accelerates the absorption of calcium from the gastrointestinal tract, as shown in Figure 327. This occurs even when calcium is in the intestine in the form of calcium phosphate, which is its usual form, such as in milk. Presumably vitamin D splits calcium away from the phosphate ion so that it can be absorbed independently of phosphate.

Even though the action of vitamin D seems to be principally on the absorption of calcium, when it is absorbed the phosphate is almost always left in a soluble form, and is then easily absorbed. For this reason vitamin D secondarily enhances the absorption of phosphate as well as that of calcium.

A vitamin D intake several hundred times that required for normal intestinal absorption of calcium causes bone absorption throughout the body, thereby releasing large quantities of both calcium and phosphate into the extracellular fluids. Vitamin D is actually any one of several similar steroid chemical compounds. Some of these have a more absorptive effect on bone than others, while still others affect mainly the absorption of calcium from the gut.

Source of vitamin D. Vitamin D is a fatty substance having a steroid structure, and it is usually formed by ultraviolet irradiation of various steroids in animal fat. For example, steroids in the

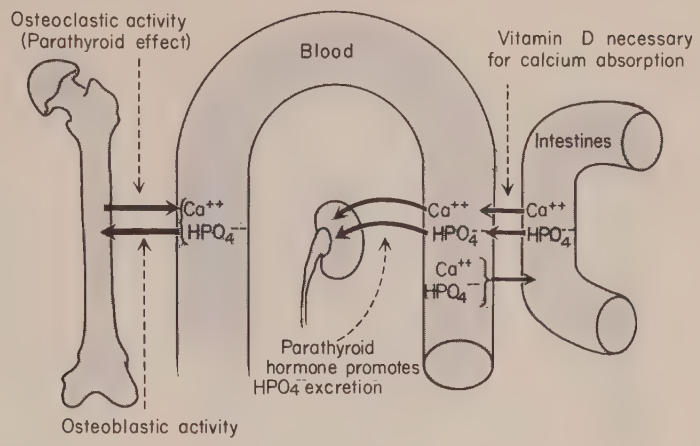


Figure 327. Absorption, utilization, and excretion of calcium and phosphate.

skin, when irradiated, form vitamin D. Consequently, a person sufficiently exposed to sunlight needs no vitamin D in his diet. If he does not receive sufficient exposure to sunlight, then generally vitamin D must be provided in his food. Usually only foods of animal origin contain vitamin D. A cow exposed to sunlight forms vitamin D continually in the skin, and some of it is secreted into the milk. Also, most animals store vitamin D in great quantities in the liver; therefore, liver is usually an excellent source of this vitamin. An artificial source of vitamin D is irradiated milk, for milk contains small quantities of steroids which can be converted into vitamin D by irradiation with ultraviolet light.

Rickets. Rickets is a disease caused by prolonged calcium deficiency. Lack of dietary calcium for a short time will never cause rickets, for when the calcium ion concentration in the extracellular fluid falls below normal, calcium is automatically absorbed from the bones, thereby reestablishing an adequate calcium ion concentration. However, if insufficient calcium is absorbed from the gut for many months, then all or most of the calcium in the bones will have been absorbed, and no more is available. The calcium ion concentration of the extracellular fluids then falls to very low values. At this point the person develops rickets, the two major effects of which are (1) depletion of calcium salts from the bones and, therefore, weakening of the bones, and (2) tetany in the late stages of the disease because of diminished extracellular fluid calcium ion concentration.

The most frequent cause of rickets is a deficiency of vitamin D rather than lack of calcium in the diet. The usual child living in temperate climates receives far too little sunlight during the winter months to provide the amount of vitamin D needed for absorption of calcium from the gut. Fortunately, the stores of vitamin D in the child's liver are usually sufficient to provide adequate calcium absorption for the early months of winter. Therefore, rickets usually develops in the early spring, before the child has been exposed to the sun again.

BONE AND ITS FORMATION

Figure 328 illustrates the general structure of a long bone, showing an articular surface at each end where it is jointed with other bones, and a hollow shaft that is excellent for resisting mechanical stresses. To the right in the figure is shown a greatly magnified cross-section of bone; the white portion of the figure represents deposits of bone salts while the black areas represent spaces that contain blood vessels and tissue fluids. Bone, like other tissue, is continually supplied with an adequate flow of nutrients from the interstitial fluid.

Chemical composition of bone. Bone is composed of a strong protein *matrix* having a consistency almost like that of leather and of *salts* deposited in this matrix to make it hard and nonbendable. By far the major portion of the salts of bone have the following chemical composition:

$$[Ca_3(PO_4)_2]_3 \cdot Ca(OH)_2$$

This is the chemical formula for hydroxyapatite, a hard marble-like compound. Small amounts of calcium carbonate ($CaCO_3$) are also present but probably are not of primary importance. The protein matrix prevents the bone from breaking when tension is applied, and the salts prevent the bone from crushing when pressure is applied. The matrix, therefore, is analogous to steel in reinforced concrete structures, and the salts are analogous to the concrete itself.

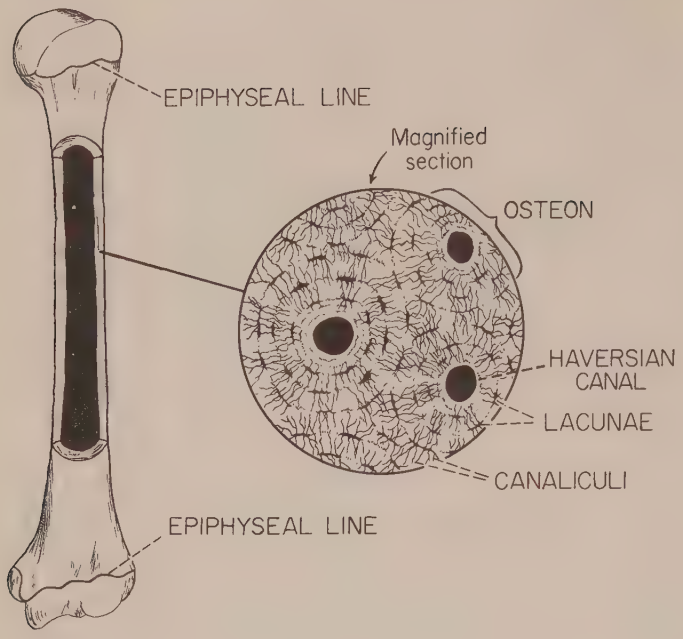


Figure 328. Structure of bone.

Deposition of Bone

Figure 329 shows, on the upper surface of a section of bone, and also in some of the cavities, a type of cell called *osteoblasts*. These are the cells that deposit bone. Bone deposition occurs in two stages as follows:

First, the osteoblasts secrete a protein substance that polymerizes to become very strong *collagen fibers*. These constitute by far the major portion of the matrix for the new bone.

Second, after the protein matrix has been deposited, calcium salts precipitate in the matrix, making it the hard structure that we know to be bone. Deposition of the salts requires (1) combination of calcium and phosphate to form calcium phosphate, $CaHPO_4$, and (2) slow conversion of this compound into hydroxyapatite over a period of several more weeks. However, the concentrations of calcium and phosphate in the extracellular fluids are normally not sufficient to cause automatic precipitation of calcium phosphate crystals. It is believed that the newly formed collagen fibers of the bone matrix have a special affinity for calcium phosphate, causing calcium phosphate crystals to deposit even though the product of calcium ion and phosphate ion concentrations is not as great as the solubility product.

Regulation of bone deposition. The deposition of bone is regulated partially by the amount of strain applied to bone. That is, the greater the weight applied and the greater the bending of the bone the more active are the osteoblasts. As a result, a bone subjected to continuous and excessive loads usually grows thick and strong, while bones not used at all, such as the bones of a leg in a plaster cast, waste away.

Another important cause of bone production is a break in a bone. Injured osteoblasts at the site of the break become extremely active and proliferate in all directions, secreting large quantities of protein matrix to cause deposition of new bone. As a result, the break is normally repaired within a few weeks.

Reabsorption of Bone

Figure 329 shows, in addition to osteoblasts, several giant cells, called *osteoclasts*, each of which contains many nuclei. These cells are present in almost all of the cavities of the bone, and have the ability to cause bone reabsorption. They do this probably by secreting an enzyme that digests the protein matrix. During the process of absorbing the matrix the salts of the bone become exposed to the extracellular fluid and are absorbed. Thus, as a result of osteoclastic activity, both calcium and phosphate are released into the extracellular fluid.

Balance between osteoclastic reabsorption and osteoblastic deposition of bone. Osteoclastic reabsorption of bone occurs all the time, but this is offset by continued osteoblastic deposition of new bone. The strength of the bone depends on the relative rates of the two processes. If the rate

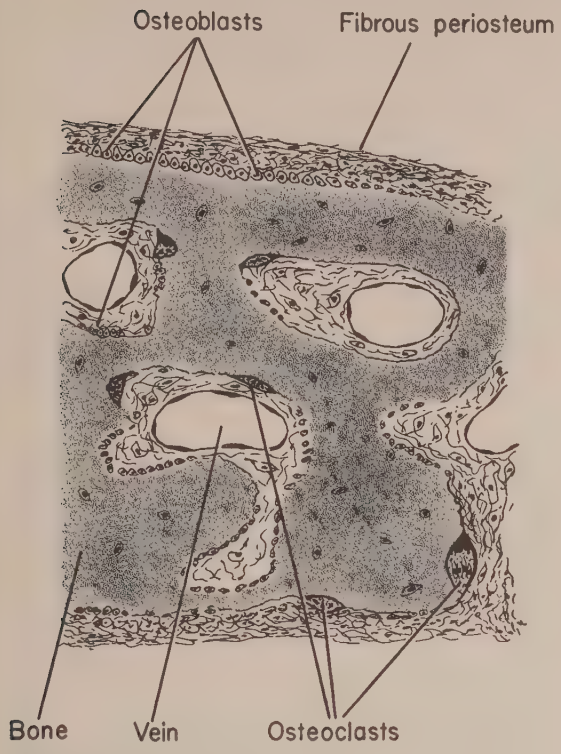


Figure 329. Osteoblastic deposition of bone, and osteoclastic reabsorption of bone.

of osteoblastic activity is greater than that of osteoclastic activity, the bone will be increasing in strength. This occurs in athletes and in others who subject their bones to excessive strain. On the other hand, osteoclastic activity is usually greater than osteoblastic activity when the bones are out of use, thereby causing the bones to become weakened.

An interesting effect of continual absorption and redeposition of bone is the tendency for crooked bones to become straight over a period of years. Compression of the inner curvature of a bent bone seems to promote rapid deposition of new bone, while stretching of the outer curvature seems to promote reabsorption. As a result, the inside of the curvature becomes filled with new bone while the outside is absorbed. In a child a broken bone may initially heal with many degrees of angulation, and yet will become essentially straight within a few years.

PARATHYROID HORMONE AND ITS REGULATION OF CALCIUM METABOLISM

Secretion by the parathyroid glands. Parathyroid hormone, also called *parathormone,* secreted by the parathyroid glands, causes release of calcium from bones, and is utilized by the body to regulate the concentration of ionic calcium in the extracellular fluids. Normally, four different parathyroid glands lying behind the four poles of the thyroid gland secrete parathyroid hormone. Each of these glands weighs only 0.02 grams.

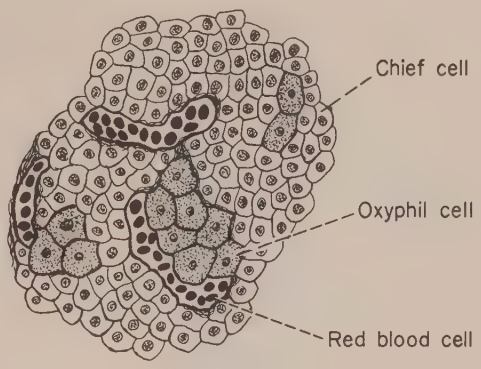


Figure 330. Histologic structure of a parathyroid gland.

A histologic cross-section of a parathyroid gland is shown in Figure 330, illustrating two types of cells in the gland, *chief cells* and *oxyphil cells.* The parathyroid glands of some animals contain only chief cells, for which reason it is assumed that the chief cells are the ones that secrete parathyroid hormone. The oxyphil cells may be the same as the chief cells but in a different state of activity.

Effect of Parathyroid Hormone on Calcium and Phosphate Ion Concentrations in the Extracellular Fluid

Administration of parathyroid hormone causes bone to be absorbed, and simultaneously both calcium and phosphate ions are dumped into the extracellular fluid. The phosphate ions are almost immediately excreted into the urine, so that very little if any change occurs in the phosphate concentration; on the other hand, calcium is slow to be excreted, and its concentration in the extracellular fluid increases considerably. Sometimes, for reasons not clearly understood, the kidney excretion of phosphate is greater than the rate of phosphate absorption from the bone. Because of this, many physiologists believe that a second important function of parathyroid hormone, in addition to its ability to absorb bone, is to promote kidney excretion of phosphate.

Basic mechanism of bone absorption caused by parathyroid hormone. Parathyroid hormone increases both the number and the sizes of the

osteoclastic cells in bone. Furthermore, after administration of parathyroid hormone, histologic study of the bone shows rapid invasion by the osteoclasts. When extreme quantities of parathyroid hormone are available, the osteoclasts become so plentiful and so large that they actually eat away enough bone to make the bones extremely weak.

Therefore, it is almost certain that parathyroid hormone removes calcium and phosphate from the bones simply by promoting osteoclastic activity.

Regulation of Parathyroid Secretion

Almost any factor that decreases the calcium ion concentration of the extracellular fluid causes the parathyroid gland to enlarge and to secrete increased quantities of parathyroid hormone. For example, a diet low in calcium increases the activity of the gland. Lack of vitamin D causes the same effect, so that the parathyroid glands frequently become four to five times normal size in persons with rickets.

In summary, parathyroid secretion responds directly to the concentration of calcium ion in the extracellular fluids. When the ionic level falls too low, the hormone is secreted, and calcium ions are absorbed into the blood from the bones, thereby replenishing the level of calcium ion. In this way, a relatively constant concentration of calcium ions is maintained in the extracellular fluids at all times.

The value of parathyroid hormone to the human being. Some of the other functions of calcium ions besides that of bone formation are actually essential to life from minute to minute. For instance, if the calcium ion concentration falls more than 50 per cent below normal, the person develops immediate tetany and dies because of spasm of his respiratory muscles. On the other hand, if the calcium ion concentration becomes too great, he is likely to have rather severe mental or cardiac disturbances. Therefore, it is necessary that the calcium ion concentration remain almost exactly constant all the time. On the other hand, the bones have almost a thousand times the 1.5 gm. of calcium in all the extracellular fluids, so that the slight amount of calcium mobilized from the bones usually is not missed. The main value of parathyroid hormone, then, is to regulate calcium ion concentration in the extracellular fluid, so that the other functions of this ion besides that of bone deposition can continue normally all the time. The bone, on the other hand, can wait several months if necessary until adequate calcium is available to replenish its lost stores.

Abnormalities of Parathyroid Hormone Secretion

Hypersecretion of parathyroid hormone. Occasionally a parathyroid tumor develops in one of the glands, and extreme secretion of parathyroid hormone causes tremendous overgrowth of the osteoclasts. Sometimes these cells grow so large and so numerous that they cause large honeycomb cavities in the bone, and even combine together to form large masses that resemble tumors. The result is often bones so weakened that very little stress can cause them to break. Indeed, most hyperparathyroid persons first become aware of their disease through a broken bone.

Hypersecretion also increases the calcium ion concentration in the body fluids, but, if the hypersecretion is not too great, the level of calcium ion will not rise enough to cause untoward effects. Rarely, though, the level becomes great enough to cause precipitation of calcium phosphate in tissues other than the bone, such as the lungs, muscles, and heart, sometimes leading to death within a few days.

Hyposecretion of parathyroid hormone. The most common cause of deficient parathyroid secretion is surgical removal of all the parathyroid glands, this usually occurring inadvertently when a surgeon is removing the thyroid gland, because of the close proximity of the parathyroid glands to this other gland.

Loss of parathyroid secretion allows the calcium ion concentration to fall so low that tetany occurs within about three days, and unless the person is treated he dies almost immediately. However, sufficient calcium to restore normal function can be mobilized from the bone within a few hours by administration of parathyroid hormone or certain of the vitamin D compounds.

A rare person has hereditary hyposecretion of parathyroid hormone. Usually the degree of hyposecretion is not sufficient to cause tetany, but it may be sufficient to cause chronically depressed osteoclastic activity in the bones. This results in brittle bones, probably because without osteoclastic activity the constant absorption and reformation of new bone ceases, and the protein matrix becomes old and brittle and is not replaced as often as needed to maintain adequate bone strength.

PHYSIOLOGY OF TEETH

The teeth cut, grind, and mix food. To perform these functions the jaws have extremely powerful muscles capable of providing an occlusive force of as much as 50 to 100 pounds between the front teeth, and as much as 150 to 200 pounds between the jaw teeth. Also, the upper and lower teeth are provided with projections and facets that interdigitate so that each set of teeth fits with the other. This fitting is called *occlusion,* and it allows even small particles of food to be caught and ground between the tooth surfaces.

Function of the Different Parts of Teeth

Figure 331 illustrates a *sagittal* section of a tooth, showing its major functional parts, *enamel, dentine, cementum,* and *pulp.* The tooth can also be divided into the *crown,* which is the portion that protrudes out of the gum into the mouth, and the *root,* which is the portion that protrudes into the bony socket of the jaw. The collar between the crown and the root, where the tooth is surrounded by the gum, is called the *neck.*

Dentine. The main body of the tooth is composed of dentine, which has a very strong bony structure. Dentine is made up principally of calcium salts of phosphate and carbonate embedded in a strong meshwork of *collagen fibers.* In other words, the principal constituents of dentine are very much the same as those of bone. The major difference is its structure, for dentine does not contain any osteoblasts, osteoclasts, nor spaces for blood vessels or nerves. Instead, it is deposited and nourished by a layer of cells called *odontoblasts,* which line its inner surface along the wall of the pulp cavity.

The calcium salts in dentine make it extremely resistant to compressional forces, while the collagen fibers make it tough and resistant to tensional forces that might result when the teeth are struck by solid objects.

Enamel. The outer surface of the tooth is covered by a layer of enamel that is formed prior to eruption of the tooth by special epithelial cells called *ameloblasts.* Once the tooth has erupted no more enamel is formed. Enamel is composed of very small crystals of calcium phosphate and calcium carbonate embedded in a fine meshwork of *keratin* fibers. The smallness of the crystalline structure of the calcium compounds makes the enamel extremely hard, much harder than dentine. Also, the keratin meshwork makes enamel very resistant to acids, enzymes, and other corrosive agents, because keratin itself is one of the most insoluble and resistant proteins known.

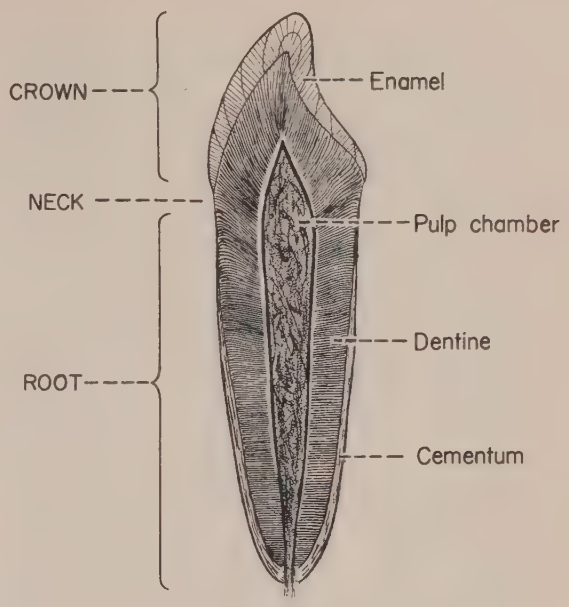


Figure 331. Functional parts of a tooth.

Cementum. Cementum is a bony substance secreted by cells that line the tooth socket. It forms a thin layer between the tooth and the inner surface of the socket, which itself is lined by an osteoblastic membrane called the *periodontal membrane.* Many collagen fibers pass directly from the bone of the jaw, through the periodontal membrane, and into the cementum. It is these collagen fibers and the cementum that hold the tooth in place. When the teeth are exposed to excessive strain the layer of cementum increases in thickness and strength. It also increases in thickness and strength with age, causing the teeth to become progressively more firmly seated in the jaws as one reaches adulthood and beyond.

Pulp. The inside of each tooth is filled with pulp, which is composed of connective tissue with an abundant supply of nerves, blood vessels, and lymphatics. The cells lining the surface of the pulp cavity are the odontoblasts, which, during the formative years of the tooth, lay down the dentine but at the same time encroach more and more on the pulp cavity, making it smaller. In later life the dentine stops growing and the pulp cavity remains essentially constant in size. However, the odontoblasts are still viable and send projections into small *dentinal tubules* that penetrate all the way through the dentine; these probably are of importance for providing nutrition.

The nerves entering the pulp do not pass into the dentine. Nevertheless, dentine that has had

its enamel layer broken away is very pain-sensitive, indicating that the projections of the odontoblasts into the dentinal tubules might in some way transmit sensations, particularly pain sensations, from the substance of the dentine into the nerve fibers of the pulp cavity.

Dentition

Human beings and most other mammals develop two sets of teeth during a lifetime. The first teeth are called the *deciduous teeth* or *milk teeth,* and they number 20 in the human being. These erupt between the seventh month and second year of life and last until the sixth to the thirteenth year. After each deciduous tooth is lost, a permanent tooth replaces it, and an additional 8 to 12 molars appear posteriorly in the jaw, making the total number of permanent teeth 28 to 32, depending on whether the person finally grows his 4 *wisdom teeth,* which do not occur in everyone.

Formation of teeth. Figure 332 shows the formation and eruption of teeth. Figure 332A shows protrusion of the oral epithelium into the *dental lamina,* this to be followed by the development of a tooth-producing organ. The epithelial cells above form ameloblasts which secrete the enamel on the outside of the tooth. The epithelial cells below extend upward to form a pulp cavity and also to form the odontoblasts that secrete dentine. Thus, enamel is secreted on the outside of the tooth, and dentine is secreted on the inside, forming an early tooth as illustrated in Figure 332B.

ERUPTION OF TEETH. During early childhood, the teeth begin to protrude upward from the jawbone through the oral epithelium into the mouth, as in Figure 332C. The cause of eruption is unknown, though several theories have been offered. One of these assumes that an increase of the material inside the pulp cavity of the tooth causes much of it to be extruded downward through the root canal, and that this pushes the tooth upward. However, a more likely theory is that the bone surrounding the tooth hypertrophies progressively, and in so doing shoves the tooth forward toward the gum margin.

DEVELOPMENT OF PERMANENT TEETH. During embryonic life a tooth-forming organ also develops in the dental lamina for each permanent tooth that will be needed after the deciduous teeth are gone. These tooth-producing organs slowly form the permanent teeth throughout the first 6 to 20 years of life. When each permanent tooth becomes fully formed, it, like the deciduous tooth, pushes upward through the bone of the jaw toward the gingival margin. In so doing it erodes the root of the deciduous tooth and eventually causes it to loosen and fall out. Soon thereafter the permanent tooth erupts to take the place of the original one.

METABOLIC FACTORS. The rate of development and the speed of eruption of teeth can be accelerated by both thyroid and growth hormones. Also, the deposition of salts in the early forming teeth is affected considerably by various factors of metabolism, such as the availability of calcium and phosphate in the diet, the amount of vitamin D present, and the rate of parathyroid hormone secretion. When all these factors are normal, the dentine and enamel will be healthy. When these factors are deficient, the ossification

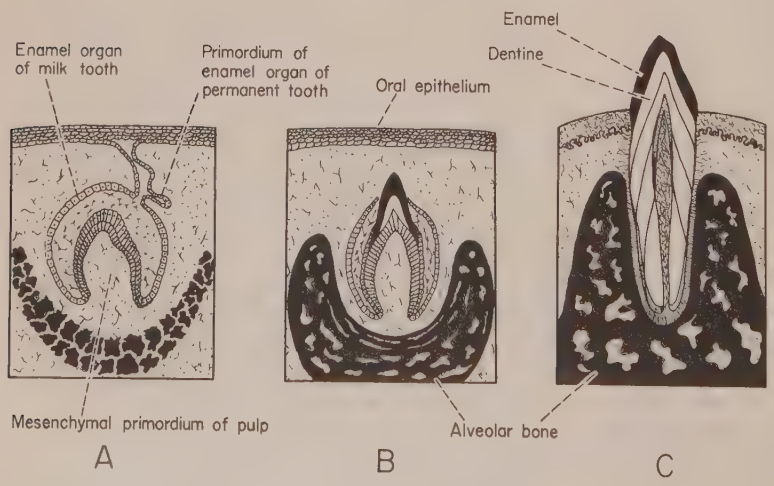


Figure 332. (A) Primordial tooth organ. (B) The developing tooth. (C) The erupting tooth. (Modified from Bloom and Fawcett: *A Textbook of Histology.* 8th Ed.)

of the teeth also may be defective, so that the teeth will be abnormal throughout life.

Mineral Exchange in Teeth

The salts of teeth, like those of bone, are composed of calcium phosphate and calcium carbonate salts bound together in a single very hard crystalline substance. New calcium salts are constantly being deposited, while old salts are being reabsorbed from the teeth, as also occurs in bone. However, experiments indicate that this deposition and reabsorption occurs in the dentine and cementum, while almost none, if any at all, occurs in the enamel. The rate of absorption and deposition of minerals in the cementum is approximately equal to that in the surrounding bone of the jaw, while the rate of deposition and absorption of minerals in the dentine is approximately one-third that of bone. The characteristics of cementum, including the presence of osteoblasts and osteoclasts, are almost identical with those of usual bone, while dentine does not have these characteristics, as was explained above; this difference undoubtedly explains the different rates of mineral exchange.

The mechanism by which minerals are deposited in and reabsorbed from the dentine is unknown. It is possible that the small processes of the odontoblasts protruding into the tubules of the dentine are capable of absorbing salts and then providing new salts to take the place of the old. It is also possible, though this too has never been proved, that the odontoblasts provide continuous replacement of the collagen fibers of the dentine. This would be comparable to the rejuvenation of the bone matrix by osteoblasts, which is necessary for adequate maintenance of bone strength.

In summary, mineral exchange is known to occur in the dentine and cementum of teeth, though the mechanism of this exchange in dentine is unknown. On the other hand, enamel seems to maintain its original mineral complement throughout life, or until some abrasive or eroding action causes its destruction.

Dental Abnormalities

The two most common dental abnormalities are *caries* and *malocclusion.* Caries means erosions of the teeth, while malocclusion means failure of the projections of the upper and lower teeth to interdigitate properly.

Malocclusion. Malocclusion is usually the result of a hereditary abnormality that causes the teeth of one jaw to grow in an abnormal direction. When malocclusion occurs, the teeth cannot perform adequately their normal grinding or cutting action. Occasionally malocclusion results in abnormal displacement of the lower jaw in relation to the upper jaw, causing such undesirable effects as pressure on the anterior portion of the ear or pain in the mandibular joint.

The orthodontist can often correct malocclusion by applying long-term gentle pressure against the teeth with appropriate braces. The gentle pressure causes absorption of alveolar jaw bone on the compressed side of the tooth, and deposition of new bone on the tensional side of the tooth. In this way the tooth gradually moves to a new position as directed by the applied pressure.

Caries. Two major but differing theories have been proposed to explain the cause of caries. One of these postulates that acids formed in crevices of the teeth by acid-producing bacteria cause erosion and absorption of the protein matrix of the enamel and dentine. This theory assumes that the acids are formed by splitting carbohydrates into lactic acid. For this reason it has been taught that eating food high in carbohydrate content, and particularly eating large quantities of sweets between meals, can lead to excessive development of caries.

The second theory proposes that proteolytic (protein-digesting) enzymes secreted by bacteria in the crevices of the teeth, or in plaques that develop on the surfaces of unbrushed teeth, digest the keratin matrix of the enamel. Then the calcium salts, unprotected by their protein fibers, are slowly dissolved by the saliva.

Which of the above theories correctly describes the primary process for development of caries is not known. Indeed, both processes might be operative simultaneously to eventuate in carious teeth.

Some teeth are more resistant to caries than others. Studies in recent years indicate that teeth formed in children who drink water containing small amounts of fluorine develop enamel that is more resistant to caries than that of the teeth of children who drink water containing no fluorine. Fluorine does not make the enamel harder than usual, but is believed to inactivate proteolytic enzymes before they can digest the protein matrix of the enamel. Regardless of the precise means by which fluorine protects the teeth, it is known that small amounts of it deposited in enamel make teeth about twice as resistant to caries as are teeth without fluorine.

REFERENCES

Albright, F., and Reifenstein, E. C., Jr.: *The Parathyroid Glands and Metabolic Bone Disease,* Selected Studies. Baltimore, The Williams & Wilkins Company, 1948.

Bartter, F. C.: The parathyroids. *Ann. Rev. Physiol.* 16:429, 1954.

Black, B. M.: *Hyperparathyroidism.* Springfield, Ill., Charles C Thomas, 1953.

Ficarra, B. J.: *Diseases of the Thyroid and Parathyroid Glands.* New York, Grune & Stratton, Inc., 1958.

Jenkins, G. M.: *The Physiology of the Mouth.* 2nd Ed. Oxford, Blackwell Scientific Publications, Ltd., 1960.

McLean, F. C., and Budy, A. M.: Connective and supporting tissues: bone. *Ann. Rev. Physiol.* 21:69, 1959.

McLean, F. C., and Urist, M. R.: *Bone.* Chicago, The University of Chicago Press, 1955.

Munson, P. L., Hirsch, P. F., and Tashjian, A. H., Jr.: Parathyroid gland. *Ann. Rev. Physiol.* 25:325, 1963.

Nicolaysen, R., Eeg-Larson, N., and Malm, O. J.: Physiology of calcium metabolism. *Physiol. Rev.* 33:424, 1953.

Rodahl, K., editor, Nicholson, J. T., and Brown, E. M.: *Bone as a Tissue.* New York, McGraw-Hill Book Company, Inc., 1960.

SEXUAL FUNCTIONS OF THE MALE AND FEMALE, AND THE SEX HORMONES

36

The male and female play equal parts in initiating reproduction and in determining the hereditary characteristics of the baby—the male provides the *sperm* and the female the *ovum.* Combination of a single sperm with a single ovum forms a *fertilized ovum* that can grow into an *embryo,* then a *fetus,* and eventually a newborn baby. The sexual functions of both the male and female responsible for initiating the process of reproduction are discussed in the present chapter, and reproduction itself is considered in the following chapter.

MALE SEXUAL FUNCTIONS

Figure 333 illustrates the male sexual organs, the most important of which are the testes, the prostate, and the penis. The functions of these are described in subsequent sections of the chapter.

The Sperm

Characteristics of sperm. The *testes* of the male produce billions of *spermatozoa,* also called *sperm.* Figure 334 shows the structure of a sperm, which itself is a single cell constructed principally of a *head* and a *tail.* The head is composed mainly

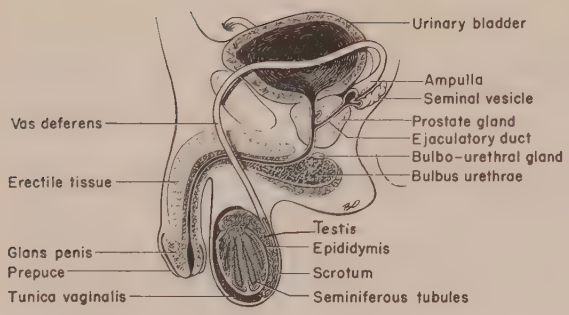


Figure 333. The male sexual organs. (Modified from Bloom and Fawcett: *A Textbook of Histology.* 8th Ed.)

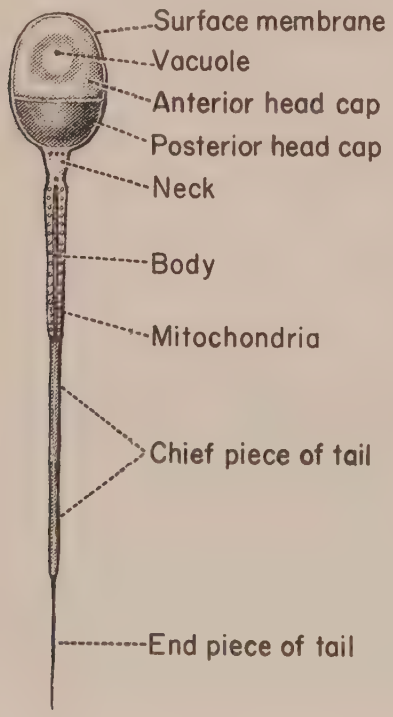


Figure 334. Structure of a sperm.

of the nucleus of the cell, with only a very thin cytoplasmic and membrane layer around its surface. Most of the cytoplasm of the sperm is aggregated into the tail, and this is covered by a long extension of the cellular membrane.

The nucleus in the head of the sperm carries the genes of the male, while the tail provides motility. As long as the sperm remains alive the tail moves back and forth *(flagellar movement)*, propelling the sperm forward. Normal sperm move in a straight line at a velocity of about 1 to 4 mm. per minute. This movement allows them to pass up the female genital tract in quest of the ovum.

Formation of sperm—spermatogenesis. The testes are composed of many thousands of small tubules, called *seminiferous tubules,* which form the sperm. A cross-section of a typical seminiferous tubule is illustrated in Figure 335A, and the formation of sperm, which is called *spermatogenesis,* is shown in Figure 335B. The epithelioid cells lining the outer wall of the seminiferous tubules are called the *male germinal epithelium,* and it is from these that all sperm are formed.

The earliest form of cell in the development of spermatozoa is the *spermatogonium.* As this cell divides through several generations, the newly forming cells move toward the lumen of the tubule. The first stage in the development of the sperm is the *primary spermatocyte;* this divides again to form the *secondary spermatocyte,* which in turn divides to form two *spermatids.* The spermatid changes into a sperm by, first, losing some of its cytoplasm, second, reorganizing the chromatin material of its nucleus to form a compact head, and, third, collecting the remaining cytoplasm at one end of the cell to form the tail.

Located among the germinal cells of the tubule are many large cells called *sustentacular cells.* These lie adjacent to the developing spermatocytes, and the spermatocytes remain attached to the sustentacular cells until the head and tail of the sperm are formed. It is believed that the sustentacular cells secrete substances that are needed by the developing sperm, but their exact function is not known.

DIVISION OF CHROMOSOMES DURING SPERM FORMATION. When a secondary spermatocyte divides to form two spermatids, the cell division is not of the usual type, for reproduction of the chromosomes does not occur. Instead, each of the 23 pairs of chromosomes simply split apart, allowing 23 unpaired chromosomes to pass into one of the two spermatids and the other 23 to pass into the other spermatid. In other words, only half of the genes enter each spermatid, and, similarly, only half of the genes from the parent cell are in each respective sperm. A similar division of genes occurs in the ovum. Thus, each parent makes an equal contribution to the hereditary characteristics of the child.

SEX DETERMINATION BY THE SPERM. Whether a baby will be male or female is determined by the sperm, for the following reasons. One pair of the chromosomes in all human cells are called the sex pair because these chromosomes determine whether the person will be male or female. One chromosome of the pair is known as the *Y* chromosome or the *male chromosome.* The other is known as the *X* chromosome or the *female chromosome.* A female has two X chromosomes.

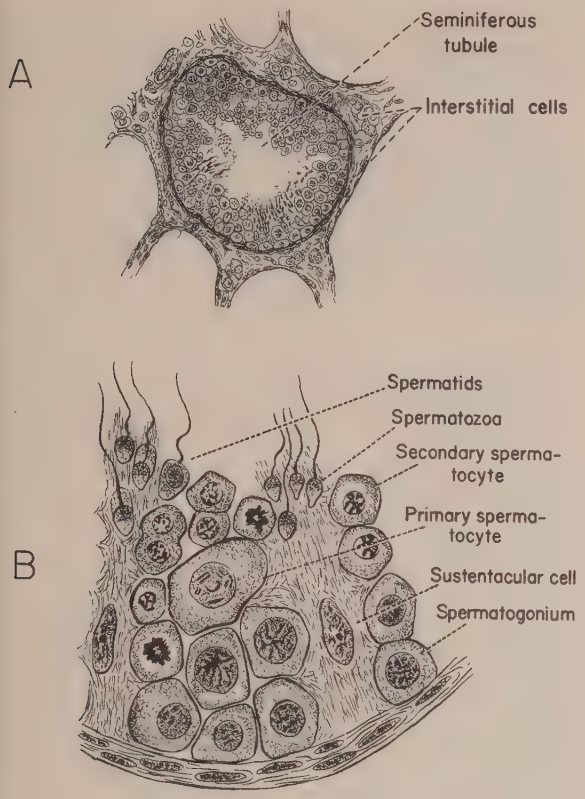


Figure 335. (A) Cross-section of a seminiferous tubule. (B) Development of spermatozoa by the germinal epithelium. (Modified from Arey: *Developmental Anatomy.* 6th Ed.)

A male has one X chromosome and one Y chromosome. When the spermatocyte divides in the testis to form sperm, the sex chromosomes split; the Y chromosome passes into one sperm and the X chromosome passes into another. Since the unfertilized ovum contains only a single X chromosome, if a sperm containing an X chromosome combines with an ovum, the fertilized ovum then has a pair of X chromosomes. This combination leads to formation of a female. On the other hand, combination of a sperm containing a Y chromosome with an ovum creates an X-Y pattern in the fertilized ovum and causes the development of a male child.

Storage of sperm. Sperm are formed continually by the seminiferous tubules. After formation they are stored mainly in the lumen of the tubules, but some of them pass along the male genital ducts, shown in Figure 333, to be stored in the *epididymis.* The sperm remain almost totally inactive during this period of storage because they release large quantities of carbon dioxide and other acidic end-products that severely depress their motility. Nevertheless, they can live for as long as a month in the seminiferous tubules and epididymis, awaiting ejaculation.

The Male Sexual Act

In addition to the testes, the male sexual organs include, as shown in Figure 333: the *vas deferens,* which conducts the sperm from the testis to the urethra; the *seminal vesicles,* which secrete a mucoid material into the upper end of the vas deferens; the *prostate gland,* which secretes a milky fluid also into the upper end of the vas deferens; the *urethra,* which conducts the sperm from the vas deferens to the exterior; and the *penis,* which provides the passageway into the female vagina. The penis in turn has two important parts for performance of the sexual act. These are the *glans,* which is the sensitive portion that elicits sexual excitement, and the *erectile tissue,* which lies on all sides of the urethra and causes erection.

Erection and lubrication. Stimulation of the glans of the penis causes sexual sensory impulses to pass into the sacral portion of the spinal cord, and, if the person simultaneously has the psychic desire to perform the sexual act, reflex impulses return through parasympathetic nerve fibers to the genital organs. These impulses dilate the arteries that supply the erectile tissue of the penis and probably also constrict the veins. As a result, a tremendous amount of blood enters the erectile tissue under high pressure and blows it up like a balloon, which causes the penis to become greatly enlarged and to extend forward.

In addition to causing erection, the parasympathetic impulses initiate mucous secretion by the *bulbo-urethral glands* located at the upper end of the urethra, and also by many small mucous glands along the course of the urethra. The mucus is expelled and helps to lubricate the movement of the penis and, therefore, allows the vagina to massage the penis rather than to abrade it. The massaging effect creates the necessary sexual stimulus to cause ejaculation; an abrasive effect causes pain, which inhibits sexual desire and blocks completion of the sexual act.

Ejaculation. When the degree of sexual stimulation has reached a critical level, neuronal centers in the tip of the spinal cord send impulses through the sympathetic nerves to the male genital organs to initiate rhythmic peristalsis in the genital ducts. The peristalsis begins in the testes themselves and then passes upward through the epididymis, the vas deferens, the seminal vesicles, the prostate gland, and the penis itself. This process, called *ejaculation,* moves sperm all the way from the testes out the tip of the penis into the vagina.

In addition to the sperm from the testes, mucus from the seminal vesicles and a milky serous fluid from the prostate are expelled during ejaculation.

All these fluids plus the sperm are called *semen*, and the total quantity in each ejaculation is usually about 3 ml., each milliliter containing about 120 million sperm, a total of about one-third billion sperm.

The milky fluid from the prostate, which is highly alkaline, neutralizes the acidic fluid from the testes, in this way immediately stimulating the sperm to action, for sperm are immobile in acidic media but very active in alkaline media.

Male sterility. Approximately one male out of every 25 to 30 is sterile. The most frequent cause of this is previous infection in the male genital ducts, though occasionally the seminiferous tubules of the testes may have been partially or totally destroyed by mumps infection, typhus infection, x-ray irradiation, or nuclear radiation. A few males have congenitally deficient testes that are incapable of producing normal sperm. The sperm may have two tails, two heads, or other less obvious abnormalities. These cannot fertilize an ovum, and when a large number of them occur in the ejaculate, it is usually an indication that all of the sperm are abnormal even though many appear to be completely normal.

Male sterility also can occur when the number of sperm in the ejaculate falls too low, even though all the sperm are normal. Usually a male is sterile when the number falls below about 150 million sperm in a single ejaculation. It is difficult to understand why sterility should exist in this instance, since only one sperm is required to fertilize the ovum. However, it is believed that the large number of sperm are necessary to provide enzymes or other substances that help the single fertilizing sperm to reach the ovum. It has been suggested, though not proved, that sperm secrete an enzyme, called *hyaluronidase*, which disperses the granulosal cells that normally cover the surface of the ovum. This action then allows the sperm to attack the ovum. Without a sufficient number of sperm to secrete hyaluronidase, the granulosal cells supposedly block the onslaught of all sperm.

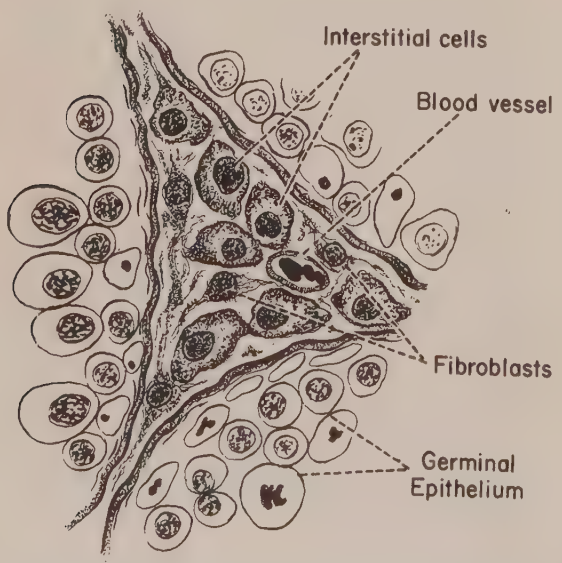


Figure 336. The interstitial cells of the testis. (Modified from Bloom and Fawcett: *A Textbook of Histology*. 8th Ed.)

HORMONAL REGULATION OF MALE SEXUAL FUNCTIONS

Role of the Adenohypophyseal Hormones

Puberty. The testes of the child remain dormant until they are stimulated at the age of 10 to 14 by gonadotropic hormones from the pituitary gland. At that age, for reasons not yet understood, the adenohypophysis begins to secrete both *follicle stimulating hormone* and *luteinizing hormone*. These cause an upsurge of testicular growth and function, causing the male sex life to begin. This stage of development is called *puberty*.

Follicle stimulating hormone. Follicle stimulating hormone causes proliferation of the cells in the germinal epithelium, promoting the formation of sperm. The means by which follicle stimulating hormone promotes spermatogenesis are not known. Yet coincident with the proliferation of the germinal cells, estrogens—female sex hormones—are produced by the sustentacular cells. Therefore, it has been suggested that follicle stimulating hormone might act primarily on the sustentacular cells to cause estrogen production, and that it may be the estrogens diffusing into the adjacent germinal cells that make them proliferate. Indeed, one of the known functions of estrogens is to cause cellular proliferation; this is discussed in detail in relation to female sex functions later in the chapter.

Luteinizing hormone. Luteinizing hormone causes the production of testosterone by the so-called *interstitial cells*, located as shown in Figure 336 in the interstices between the seminiferous tubules. Testosterone in turn is the major hormone responsible for development of male sexual characteristics.

Functions of Testosterone

Effect on spermatogenesis. Testosterone causes the testes to enlarge. Also, it must be pres-

ent along with follicle stimulating hormone before spermatogenesis will occur. Unfortunately, though, the precise action of testosterone in spermatogenesis is unknown.

Effect on male sex characteristics. After a fetus begins developing inside its mother's uterus, its testes begin to secrete testosterone when it is only a few weeks old. This testosterone then helps the fetus to develop male sexual organs and male secondary characteristics. That is, it accelerates the formation of a penis, a scrotum, a prostate, the seminal vesicles, the vas deferens, and other male sexual organs. In addition, testosterone causes the testes to descend from the abdominal cavity into the scrotum; if the production of testosterone by the fetus is insufficient, the testes fail to descend, but instead remain in the abdominal cavity in the same manner that the ovaries remain in the abdominal cavity of the female.

Testosterone secretion by the fetal testes is caused by a hormone, called *chorionic gonadotropin*, that is formed in the placenta during pregnancy. Immediately after birth of the child, loss of connection with the placenta removes this stimulatory effect, so that the testes become dormant and the sexual characteristics remain at a standstill from birth until puberty. At puberty the reinstitution of testosterone secretion causes the male sex organs to begin growing again. The testes, scrotum, and penis then enlarge about ten-fold.

Effect on secondary sex characteristics. In addition to the effects on the genital organs, testosterone exerts other general effects throughout the body to give the adult male his distinctive characteristics. It promotes *growth of hair* on his face, along the midline of his abdomen, on his pubis, and on his chest. On the other hand, it causes *baldness* in those male individuals who have a hereditary predisposition to baldness. It increases the *growth of the larynx* so that the male, after puberty, develops a deeper pitch to his voice. It causes an *increase in the deposition of protein* in his muscles, bones, skin, and other parts of his body, so that the male adolescent becomes generally larger and more muscular than the female. Also, testosterone sometimes promotes abnormal secretion by the sebaceous glands of the skin, this leading to acne on the face of the post-puberal male.

FEMALE SEXUAL FUNCTIONS

Figure 337 shows the sexual organs of the female. The general function of these is the following: An *ovary* forms an *ovum*, which is then transported through a *fallopian tube* into the *uterus*. If a sperm was available to fertilize the ovum, the fertilized ovum develops in the uterus into a fetus and then into a child. The purpose of the present section is to explain the female sexual functions that are necessary to initiate the process of reproduction.

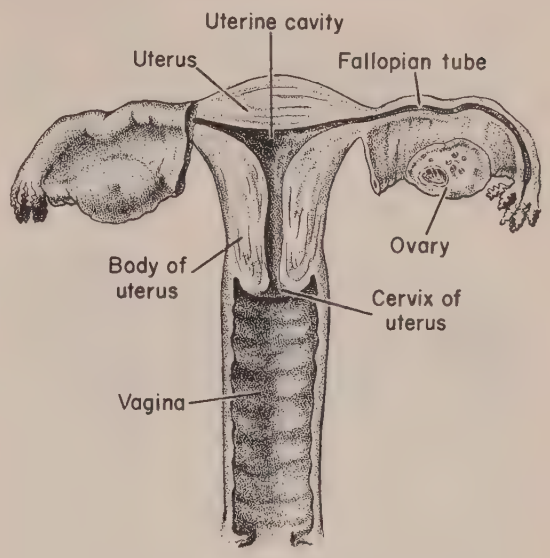


Figure 337. The female sexual organs.

The Ovum

The ovaries, like the testes, have a germinal epithelium. However, this is not arranged in tubules as in the testes, but instead lies on the surface of each ovary. In the early stages of ovarian development some of the cells from this germinal epithelium migrate into the mass of the ovary and after several divisions form large cells which become the *ova*. Approximately 400,000 ova in all are formed by the two ovaries, most of which are already present in the substance of the ovaries at birth and puberty. At this time, hormones from the pituitary gland begin promoting the release of ova one at a time, as explained below.

Each ovum in the ovary is surrounded by a layer of epithelioid cells called *granulosal cells*, and these along with the enclosed ovum are called a *primary follicle*, one of which is illustrated in Figure 338.

Growth of the follicle and discharge of the ovum. At puberty, when the ovaries begin to be stimulated by gonadotropic hormones from the adenohypophysis, a few of the primary follicles

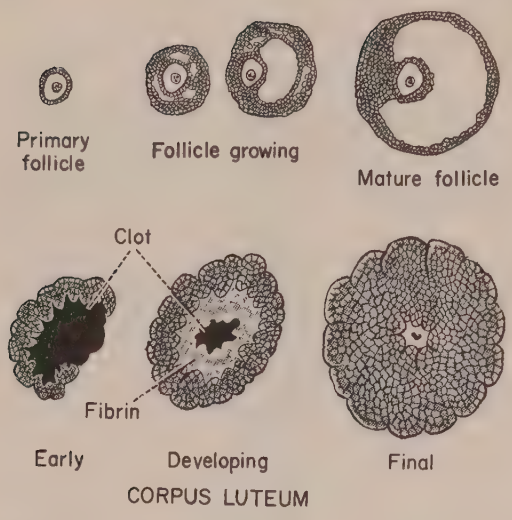


Figure 338. Growth of the follicle, and formation of the corpus luteum. (Modified from Arey: *Developmental Anatomy.* 6th Ed.)

begin to enlarge. The granulosal cells proliferate, and adjacent stromal cells from the substance of the ovary also take on epithelioid characteristics and join the growing follicle. The follicular cells then secrete fluid, creating a cavity called an *antrum* in the follicle, as shown in Figure 338. The antrum grows until the follicle actually balloons outward on the surface of the ovary; several of these follicles are shown in the ovary to the right in Figure 337.

One of the follicles soon balloons outward more than the others and ruptures, and an ovum, covered by a mass of granulosal cells, is expelled into the abdominal cavity. As soon as this follicle ruptures, the remainder of the growing follicles suddenly begin to disappear. Presumably the release of the follicular fluid into the abdominal cavity and absorption of its hormones by the peritoneum in some way cause the involution of the other follicles. Regardless of the exact cause, only one ovum is usually expelled into the abdominal cavity each month, though occasionally a second or even more ova are expelled before the remaining follicles can start to involute. This is the major cause of multiple births.

DIVISION OF CHROMOSOMES IN THE OVUM. Figure 339 shows the development of the mature ovum from the germinal epithelium. After several divisions a germinal epithelial cell becomes a *primary oocyte,* which is the ovum of a primary follicle. Toward the end of growth of the follicle the nucleus of the primary oocyte divides, but without reproducing the chromosomes. Instead, each pair of chromosomes splits apart and 23 unpaired chromosomes remain in the ovum while the other 23 chromosomes are expelled in the so-called *first polar body.* The ovum now is called a *secondary oocyte,* and it has only half the genes of the mother. Then, approximately when the follicle ruptures, the secondary oocyte divides again; this time each of the unpaired chromosomes is reproduced, and half of them pass into the *second polar body* while half remain in the ovum. The ovum at this point is said to be mature, and it is now ready to be fertilized by a sperm.

Transport of the ovum. Referring once again to Figure 337, it can be seen that the ends of the fallopian tubes lie in close proximity to the ovaries. Cilia on the epithelial surfaces of the fallopian tubes are always beating toward the uterus, causing fluids from the region of the ovary to be moving constantly out of the abdominal cavity into the uterus. The ovum, on being expelled into the abdominal cavity from the rupturing follicle, is transported into a fallopian tube along with this current of fluid flow. The epithelium of the fallopian tube, however, is pocked along its entire extent with large cavities that obstruct the movement of the ovum, so that usually three to seven days are required for its passage to the uterus.

The ovum must be fertilized within 8 to 24 hours after it is released from the ovary or it dies. Therefore, fertilization usually takes place in the upper portion of the fallopian tube, and the fertilized ovum begins to divide and to form the early stages of the embryo while it is still in the tube. It is believed that the fallopian tube epithelium

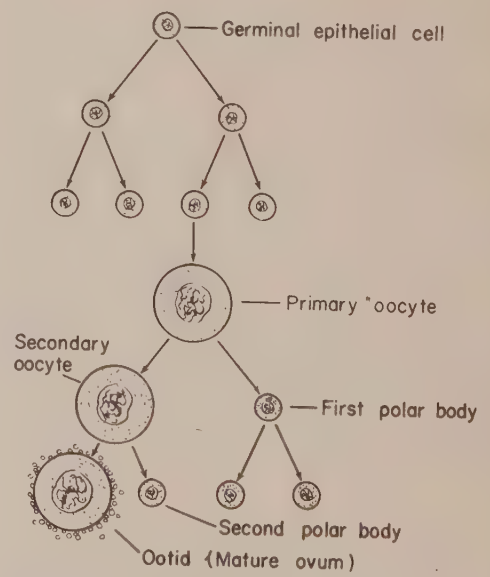


Figure 339. Formation of the mature ovum from the germinal epithelium. (Modified from Arey: *Developmental Anatomy.* 6th Ed.)

secretes substances that provide some of the nutrition for the developing mass of cells. However, much if not most of the nutrition is supplied by the large amount of cytoplasm in the ovum itself.

Female sterility. About one female out of 15 is sterile. As is true in the male, many of these instances of sterility are caused by previous infection. The infection sometimes blocks the fallopian tubes, or at other times involves the ovaries in a mass of scar tissue. The most common infection causing female sterility, as is true also in the male, is gonorrhea. Sterility is often caused also by failure of the ovaries to develop and expel ova into the abdominal cavity. Most frequently this results from insufficient stimulation by the adenohypophyseal hormones. At other times it is caused by excessively thick capsules on the surfaces of the ovaries, or occasionally by congenitally abnormal ovaries.

The Female Sexual Act

Part of the female's role in the sexual act is to help the male reach a sufficient degree of sexual stimulation for ejaculation to occur. She does this by providing appropriate conditions for maximum masculine stimulation, including especially female erection and the secretion of adequate lubricant.

Female erection and sexual lubrication. Located around the opening of the vagina are masses of erectile tissue exactly like that in the male penis. Appropriate psychic excitation in the female, plus sexual stimulation of the female genital organs, causes parasympathetic impulses to pass from the caudal spinal cord to this erectile tissue, making it swell and provide a tight though distensible opening into the vaginal canal.

In addition to erection the parasympathetic impulses cause two small glands called *Bartholin's glands*, located immediately inside the vagina on either side, to secrete large quantities of mucus. It is this mucus from the female that is mainly responsible for lubrication of the sexual act, though mucus from the male helps to a less extent.

The female climax. Most of the sexual stimulation in the female is provided by the massaging action of the penis against the *female clitoris*. It is the homologue in the female of the masculine penis, and is located at the anterior margin of the vulva immediately outside the vagina. The female, like the male, develops intense sexual stimulation, and when the degree of stimulation reaches sufficient intensity, the uterus and fallopian tubes begin rhythmic, upward peristaltic contractions. This stage is called the *female climax*. The peristaltic contractions are believed to propel semen from the vagina into the fallopian tubes.

HORMONAL REGULATION OF FEMALE SEXUAL FUNCTIONS

Relation of the Adenohypophyseal Gonadotropic Hormones to the Ovarian Cycle

The adenohypophysis of the female child, like that of the male child, secretes essentially no gonadotropic hormones until the age of 10 to 14 years. However, at that time, for reasons not at all understood, the adenohypophysis begins to secrete three gonadotropic hormones. At first it secretes only *follicle stimulating hormone*, which initiates the beginning of sexual life in the growing female child, but later it secretes *luteinizing* and *luteotropic hormones*, which help to control the monthly female cycle.

Follicle stimulating hormone. Follicle stimulating hormone causes a few of the primary follicles of the ovary to begin growing each month, promoting very rapid proliferation of the epithelioid cells surrounding the ovum. These cells then begin to secrete estrogens, one of the two major female sex hormones. Thus, the two functions of follicle stimulating hormone are to cause proliferation of the ovarian follicular cells and to cause secretory activity by these cells which causes the follicular cavities to develop and grow. As soon as the follicles grow to about one-half their maximum size, the adenohypophysis begins to secrete *luteinizing hormone* instead of follicle stimulating hormone.

Luteinizing hormone. This hormone increases still more the rate of secretion by the follicular cells, and soon makes one follicle grow so large that it ruptures, expelling its ovum into the abdominal cavity. Simultaneously, luteinizing hormone causes the follicular cells to grow in size and to develop a fatty, yellow appearance. These cells are then known as *lutein cells*, and the entire mass of lutein cells is called a *corpus luteum*, which means a "yellow body." This change of follicular cells into lutein cells is illustrated in the lower part of Figure 338.

Luteotropic hormone. The adenohypophysis also secretes a third hormone called *luteotropic hormone*, though little is known about the quantities secreted and the time in the cycle that it is secreted. This hormone further enhances the development of the corpus luteum and stimulates it to secrete large quantities of both estrogens and progesterone, the two female sex hormones. The

corpus luteum persists for about two weeks, at which time it degenerates. Then, the adenohypophysis begins secreting large quantities of follicle stimulating hormone once again, which initiates a new cycle of growing follicles.

Ovarian Hormones

The two ovarian hormones, the *estrogens* and *progesterone*, are responsible for sexual development of the female and also for the female monthly sexual cycle. These hormones, like the adrenocortical hormones and the male hormone testosterone, are both steroid compounds, and they are formed mainly from the fatty substance cholesterol. Estrogens are actually several different hormones, but they have identical functions and almost but not exactly identical chemical structures. For this reason they are considered together as if they were a single hormone.

Functions of estrogens. Estrogens cause many types of cells in certain parts of the body to *proliferate*—that is, to increase in number. For example, they cause the smooth muscle cells of the uterus to proliferate, making the female uterus, after puberty, enlarge to about two to three times that of a child. Also, estrogens cause enlargement of the vagina, development of the *labia* surrounding the vagina, growth of hair on the pubis, broadening of the hips, conversion of the pelvic outlet into an ovoid shape rather than the funnel shape of the male, growth of the breasts, proliferation of the glandular elements of the breasts, and finally deposition of fatty tissues in characteristic female areas such as on the thighs and hips. In summary, essentially all of the characteristics that distinguish the female from the male are caused by estrogens, and the basic reason for the development of these characteristics is the ability of estrogens to promote proliferation of the respective cellular elements in certain regions of the body.

Estrogens also increase the growth rate of all bones immediately after puberty, but they cause the growing portions of the bone to "burn out" within a few years, so that growth then stops. As a result, the female grows very rapidly for the first few years after puberty and then ceases growing all together. On the other hand, the male child continues to grow even beyond this time and grows to a taller height than the female—not because of more rapid growth but because of more prolonged growth.

Estrogens also have very important effects on the internal lining of the uterus, the *endometrium*, which is discussed below in relation to the menstrual cycle.

Functions of progesterone. Progesterone has little to do with the development of the female sexual characteristics; instead, it is concerned principally with preparing the uterus for acceptance of a fertilized ovum and preparing the breasts for secretion of milk. In general, progesterone increases the degree of secretory activity of the glands of the breasts and also of the cells lining the uterine wall. This latter effect is also discussed below in relation to the menstrual cycle. Finally, progesterone inhibits the contractions of the uterus, and prevents the uterus from expelling a fertilized ovum that is implanting or a fetus that is developing.

Regulation of the Female Sexual Cycle

Oscillation between the pituitary and ovarian hormones. The monthly female sexual cycle is caused by alternating secretion of follicle stimulating hormone by the adenohypophysis and estrogens by the ovary. The cycle of events that causes this alternation is the following: During the first part of the monthly cycle, the adenohypophysis secretes follicle stimulating hormone. This causes the ovaries to secrete estrogens. Then the estrogens exert an inhibitory effect on the adenohypophysis to make it stop secreting follicle stimulating hormone. Lack of this hormone stops the ovaries from secreting estrogens. Lack of estrogens removes the inhibition of the adenohypophysis and it begins secreting follicle stimulating hormone again, thus initiating a new cycle.

Before puberty the adenohypophysis is unable to secrete any gonadotropic hormones. Consequently, until that time the alternation between follicle stimulating hormone and estrogens cannot take place. As soon as a sufficient quantity of follicle stimulating hormone begins to be secreted, the sexual life of the female begins.

After about 30 years of sexual life, the ovaries "burn out" because all the primary follicles finally have grown into mature follicles and involuted. The sexual cycles cease because the ovaries no longer have any follicular cells to secrete estrogens, even though the adenohypophysis continues to secrete large quantities of follicle stimulating hormone for the remainder of the woman's life. This cessation of the monthly sexual cycles occurs in the average woman at an age of about 45, and this stage of her life is called the *menopause*.

RELATION OF LUTEINIZING HORMONE AND OF PROGESTERONE TO THE OSCILLATING CYCLE. It is quite certain that the primary controlling mechanism of the female sexual cycle is the simple oscillation between follicle stimulating hormone and estrogens, for the cycle continues even when

the other hormones are not secreted at all or in very small quantity. However, if ovulation occurs, the amount of luteinizing hormone secreted by the adenohypophysis becomes greatly increased, and this causes the ruptured follicle to grow into a corpus luteum and produce progesterone as well as still more estrogens. Figure 340 shows the secretion of the various adenohypophyseal and ovarian hormones at different stages during the monthly cycle.

The endometrial cycle. It is evident from Figure 340 that during the first half of the monthly cycle the only ovarian hormone secreted in quantity is the estrogens. Then during the latter half both estrogens and progesterone are secreted.

Estrogens cause the endometrium—that is, the lining of the inside of the uterus—to grow in thickness. Both the epithelial cells on the surface and the deeper layers of the endometrium proliferate approximately three-fold. Also, the glands of the endometrium increase greatly in depth and in tortuosity. These changes, which are called the *proliferative phase* of endometrial development, and which are illustrated in the first portion of Figure 341, occur for about the first 10 to 14 days of the monthly cycle. At approximately the fourteenth day of the cycle the corpus luteum begins secreting progesterone, which then has the following effects on the endometrium: the endometrial glands begin secreting a nutrient fluid that can be used by a fertilized ovum before it implants; large quantities of fatty substances and glycogen deposit in the deeper cells of the endometrium; and blood flow to the endometrium increases. In other words, the major function of progesterone is to make available an adequate supply of nutrients for an embryo should it begin to develop in the uterus.

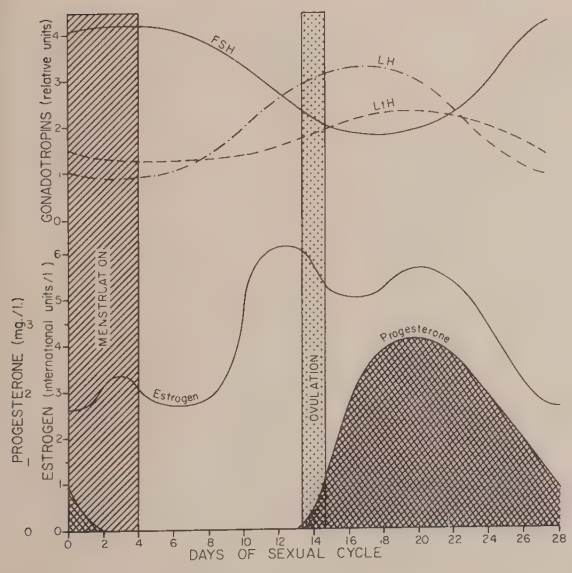


Figure 340. Secretion of the different adenohypophyseal and ovarian hormones during the normal monthly sexual cycle.

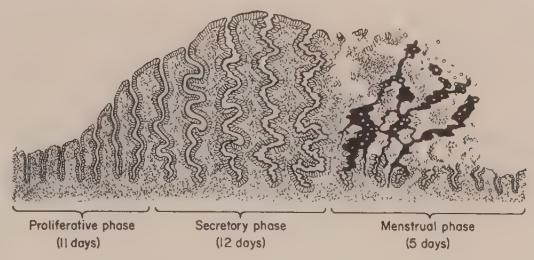


Figure 341. Endometrial changes during the monthly sexual cycle, showing especially the process of menstruation.

MENSTRUATION. If an ovum becomes fertilized and implanted in the uterus, special hormones are then released by the developing ovum itself to stimulate the corpus luteum. These cause it to continue to produce estrogens and progesterone, and pregnancy begins. However, if toward the end of the monthly cycle implantation of an ovum has not occurred, the corpus luteum dies, and the production of both estrogens and progesterone either ceases or almost ceases. The sudden lack of these two hormones causes the blood vessels of the endometrium to become spastic. The endometrial tissues die and slough into the uterine cavity, a process called *menstruation*. Blood oozes from the denuded endometrial wall, causing a blood loss of about 50 ml. during the several days of menstruation. The necrotic endometrial tissue, plus the blood and much serous exudate from the denuded uterine surface, is gradually expelled for about three to five days.

The deepest pits of the endometrial glands remain intact during menstruation, despite the sloughing of all the outer layers of the endometrium. During the first few days after menstruation begins, new epithelium grows from the edges of the glands to cover the inner surface of the uterus. Then, under the influence of renewed estrogen production by the ovaries, the endometrial cycle begins again. These sequential changes during the entire cycle are shown in Figure 341.

The period of fertility during the sexual cycle. An ovum can be fertilized by a sperm for a period of 8 to 24 hours after ovulation. Also, sperm can live in the female genital tract usually for 24 to 72 hours. Therefore, for successful fertilization, sexual exposure must occur either shortly before ovulation so that sperm are already avail-

able when ovulation occurs, or within a few hours after ovulation.

Ovulation usually occurs on the fourteenth day of the normal 28-day sexual cycle. However, many women, instead of having normal 28-day cycles, have cycles as short as 21 days or as long as 40 days, and in some women the periodicity of the cycles is very irregular. Yet despite the length of the cycle, ovulation occurs almost exactly 14 days *before* menstruation begins. Therefore, in those women who have regular sexual cycles, regardless of how long they may be, the day of ovulation can be calculated to be 14 days before the predicted day of menstruation. And because of the viability of sperm and the ovum, the period of fertility is usually between 17 days and 12 days *before* menstruation begins.

REFERENCES

Bishop, D. W.: Sperm motility. *Physiol. Rev.* 42:1, 1962.

Bullough, W. S.: *Vertebrates' Sexual Cycles.* New York, John Wiley & Sons, Inc., 1951.

Buxton, C. L., and Southam, A. S.: *Human Infertility.* New York, Paul B. Hoeber, Inc., 1958.

Crew, F. A. E.: *Sex Determination.* 3rd Ed. New York, John Wiley & Sons, Inc., 1954.

Farris, E. J.: *Human Ovulation and Fertility.* Philadelphia, J. B. Lippincott Company, 1959.

Kinsey, A. C.: *Sexual Behavior in the Human Female.* Philadelphia, W. B. Saunders Company, 1953.

Kinsey, A. C.: *Sexual Behavior in the Human Male.* Philadelphia, W. B. Saunders Company, 1948.

Mann, T. R. R.: *Biochemistry of Semen.* New York, John Wiley & Sons, Inc., 1954.

Velardo, J. T.: *Endocrinology of Reproduction.* New York, Oxford University Press, 1958.

Whitelock, O., editor: The Uterus. *Ann. N. Y. Acad. Sc.* Vol. 75, January, 1959.

REPRODUCTION AND FETAL PHYSIOLOGY

37

FERTILIZATION OF THE OVUM AND EARLY GROWTH

Entry of the sperm into the ovum. After the ovum is expelled from the ovary, it remains viable for 8 to 24 hours. During this time it usually moves approximately one quarter of the distance down the tube toward the uterus. Obviously, then, fertilization must occur either in the abdominal cavity before the ovum enters a fallopian tube, or in the upper portion of one of the tubes. Sperm travel at a velocity of only 1 to 4 mm. per minute, and the total length of a fallopian tube is approximately 15 cm. Therefore, the minimum time usually required for a sperm to reach the ovum is approximately 40 minutes.

The ovum, when released from the ovary, carries on its surface a large number of granulosal cells which collectively are called the *corona radiata,* shown in Figure 342. It is claimed, though not proved, that the corona radiata is dispersed by an enzyme called *hyaluronidase* secreted by the sperm. This enzyme, which was discussed in the previous chapter, supposedly digests the protein linkages that hold the cells together, and allows the cells to break away from the ovum. In other words, the sperm supposedly removes this barrier so that it can attack and enter the ovum.

After a single sperm has entered the ovum, the membrane of the ovum suddenly becomes impermeable to entry of additional sperm. Sperm

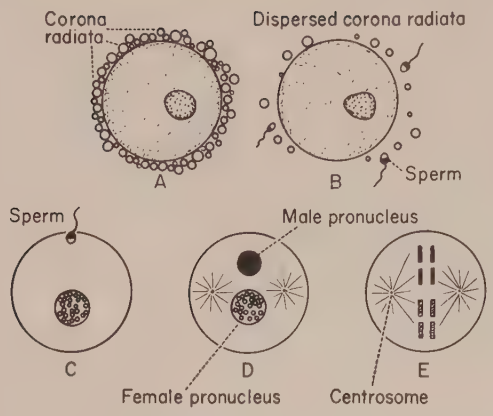


Figure 342. Fertilization of the ovum by the sperm, and beginning cleavage. (Modified from Arey: *Developmental Anatomy.* 6th Ed.)

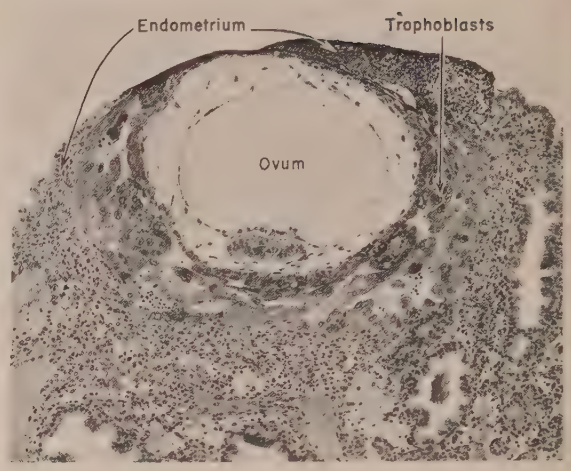


Figure 343. Implantation of a 1½ week old human ovum in the endometrium. (Courtesy of Dr. Arthur Hertig.)

that do reach the ovum are inactivated as they attempt to enter. This effect prevents more than one set of male chromosomes from combining with the chromosomes of the ovum.

Shortly after a sperm has entered the ovum the head of the sperm begins to swell, forming a *male pronucleus,* and the original nucleus of the ovum becomes the *female pronucleus.* These two nuclei soon combine so that the fertilized ovum now has 46 chromosomes, as do all other cells of the human body. A few hours later the chromosomes reproduce, become paired off, and split apart, initiating division of the ovum into two daughter cells.

Early division of the ovum. The first cleavage of the fertilized cell occurs approximately 30 hours after the sperm has entered the ovum, and successive generations occur every 10 to 15 hours. By the time the ovum reaches the uterus, the total number of cells in the cellular mass is usually about 16 to 32. By this time the cells have actually begun to *differentiate,* which means that some of them have developed different characteristics from the others.

Implantation of the ovum. Figure 343 shows the *blastocyst* stage of a developing ovum 1½ weeks old. It illustrates the method by which the mass of cells attaches itself to the inner wall of the uterus. At this stage the cells on the outside of the blastocyst, called *trophoblasts,* secrete large quantities of proteolytic enzymes that digest the endometrium, and then the trophoblasts phagocytize the digested products. Thus, they eat their way into the wall of the endometrium. The trophoblasts grow and divide very rapidly at the same time, and soon they and adjacent cells begin forming the placenta and fetal membranes while the embryo develops on the inside of the blastocyst.

NUTRITION OF THE FETUS IN THE UTERUS

The Trophoblastic Phase of Nutrition

During the first few weeks after implantation of the ovum, the placenta and its blood supply are not adequately formed to supply the fetus with nutrients. During this phase nutrition is provided only by trophoblastic digestion and phagocytosis of the endometrium. It will be recalled that prior to implantation of the ovum, the endometrium stores large quantities of proteins, lipid materials, and glycogen. Also, small quantities of iron and vitamins are stored awaiting phagocytosis by the developing ovum. Even so, the embryo can obtain nutrition in this manner for only the first few weeks of its development. By that time the placenta will have developed to a stage such that it can begin supplying nutrition. Figure 344 shows that the trophoblastic period of phagocytic nutrition lasts until approximately the twelfth week of pregnancy.

Placental Nutrition of the Fetus

Figure 345 shows the *placenta* and the *fetal membranes.* The fetal membranes attach to the entire inner surface of the uterus, forming a cavity called the *amniotic cavity.* The placenta covers about one-fourth the surface of the uterus, and the fetus is connected with it by means of the *umbilical cord.* The fetus floats freely in *amniotic fluid* that fills the amniotic cavity. The function of the

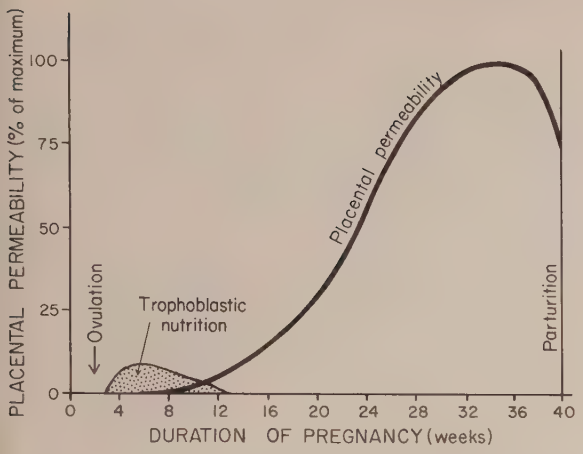


Figure 344. Periods of fetal nutrition: first, trophoblastic phagocytosis of the endometrium lasting for the first few weeks of pregnancy and, second, diffusion through the placenta during the remainder of pregnancy.

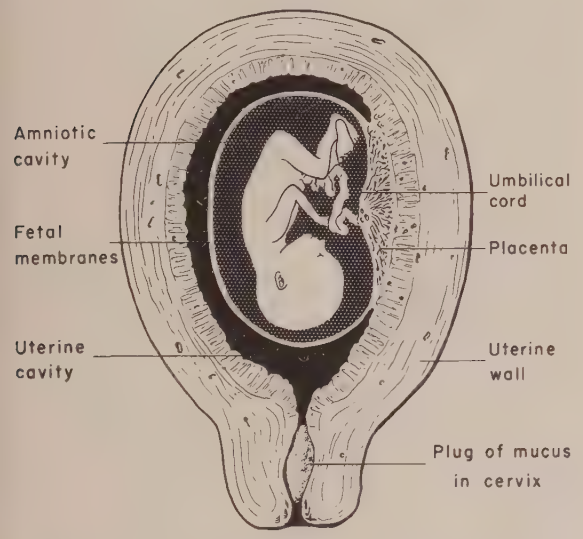


Figure 345. The fetus, the fetal membranes, and the placenta in the pregnant uterus.

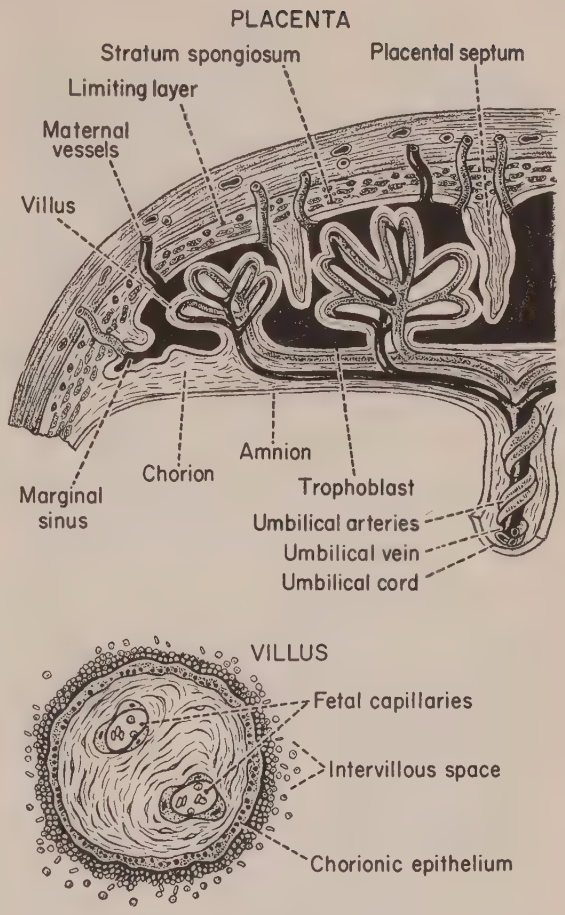


Figure 346. Gross anatomical structure of the placenta and microscopic cross-section of a placental villus. (Modified from Goss: *Gray's Anatomy of the Human Body.* Lea & Febiger; and from Arey: *Developmental Anatomy.* 6th Ed.)

umbilical cord is to transport blood from the baby to the placenta and then back again to the baby. In this way the baby's blood picks up nutrition from the placenta and transports it into the developing child.

Figure 346 illustrates the gross organization of the placenta and, at the bottom, a cross-section of a *placental villus.* The mother's blood flows into large placental chambers called *placental sinuses;* the blood passes through these so rapidly that it remains fresh essentially all the time. The fetal portion of the placenta is composed of many small cauliflower-like projections of tissue extending into the placental sinuses. Each of these in turn is covered with tremendous numbers of small villi containing blood capillaries. The baby's blood flows into the capillaries of the villi where it receives nutrients from the mother's blood, and then it passes back through the umbilical veins to the fetus. The cross-section of a villus in Figure 346 shows the close proximity of the mother's blood in the placental sinuses to the baby's blood in the fetal capillaries.

Diffusion of nutrients through the villi. Nutrients pass through the placental villi into the fetal blood almost entirely by the process of diffusion. The gaseous pressure of oxygen in the blood of the maternal sinuses is normally 60 to 80 mm. Hg, whereas the gaseous pressure of oxygen in the fetal capillaries is only 20 to 30 mm. Hg. Because of this pressure difference, oxygen simply diffuses through the tissues of the villus from the mother's blood to the baby's blood. Likewise,

glucose, some amino acids, some fats, many of the vitamins, and most of the minerals are present in greater concentration in the mother's blood than in the baby's blood because the baby utilizes these substances almost as rapidly as they enter his blood. As a consequence, all of them diffuse inward and thereby supply the fetus with nutrition.

Figure 344 shows the progressive increase in placental permeability as pregnancy proceeds. The total amount of nutrients that can be transported through the placenta reaches a maximum 32 to 36 weeks after pregnancy begins. At this point the placental tissues begin to grow old and degenerate. As a result, even though the fetus still requires progressively increasing quantities of nutrients, the ability of these nutrients to enter the baby becomes lessened. Fortunately birth of the baby occurs at this time, and the baby assumes independent existence.

EXCRETION THROUGH THE PLACENTA BY DIFFUSION. In addition to the diffusion of nutrients into the baby, excretory products diffuse through the placenta from the fetal blood to the maternal blood. For instance, the baby's metabolism continually forms large quantities of carbon dioxide, urea, uric acid, creatinine, phosphates, sulfates, and other normal excretory products. The concentration of each of these substances rises in the baby's blood until it is greater than the concentration in the mother's blood. Then, because of this reverse concentration gradient, these substances diffuse backward through the villi, and the mother excretes them through her kidneys, lungs, and other excretory organs.

Active transport through the villi. The chorionic epithelium that covers the surface of the placental villi is formed from the trophoblastic cells of the early fetus. These cells continue their phagocytic activity throughout the life of the placenta, so that some nutrients are actively transported through the villi all through the period of gestation. In the first few weeks of placental development this aids greatly in increasing the amount of amino acids, fatty substances, and some minerals that can be supplied to the baby, but after 12 to 20 weeks the amount of nutrients supplied in this manner becomes far less than that supplied by the simple process of diffusion.

THE HORMONES OF PREGNANCY

Secretion of Estrogens during Pregnancy

During the normal menstrual cycle moderate quantities of estrogens are secreted first by the follicular cells and then by the corpus luteum of the ovaries. If the ovum becomes fertilized and implants in the uterus, the corpus luteum, instead of degenerating as it normally does at the end of the normal monthly cycle, grows even larger and produces increasing amounts of estrogens for an additional four to five months. At this time, the placental tissues begin to form extreme quantities of estrogens, the amount progressively increasing to a maximum immediately before birth, as shown in Figure 347. The peak production reaches about 50 times as much as the production during a normal monthly sexual cycle.

Functions of estrogens during pregnancy. Estrogens cause rapid proliferation of the uterine

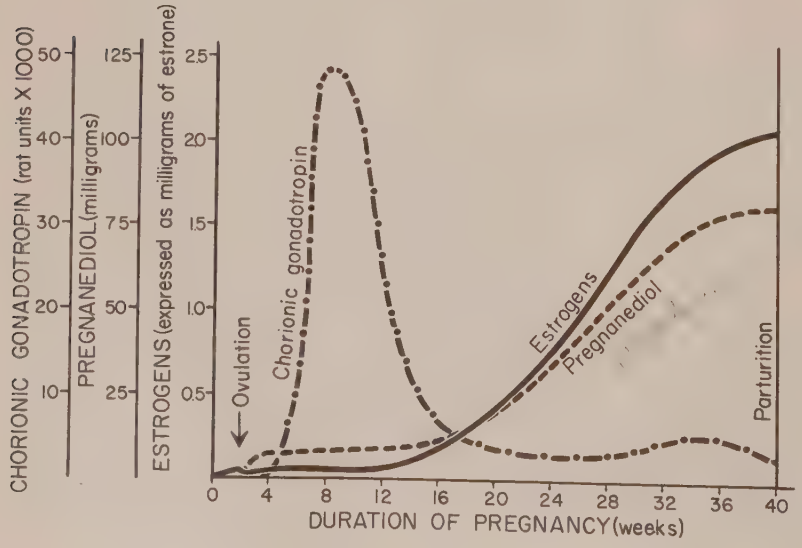


Figure 347. Secretion of hormones during the course of pregnancy.

musculature; greatly increased growth of the vascular system supplying the uterus; enlargement of the external sex organs and of the vaginal opening, providing an appropriately enlarged passageway for birth of the baby; and probably also some of the relaxation of the pelvic ligaments that allows the pelvic opening to stretch as the baby is born.

In addition to their effects on the reproductive organs, estrogens also cause the breasts to grow rapidly. The ducts, especially, enlarge, and the glandular cells proliferate. Finally, estrogens cause a pound or more of extra fat to deposit in the breasts.

Another effect of estrogens not yet completely understood is their effect on the fetus itself. It is believed that they cause much of the rapid proliferation of fetal cells and also aid in the differentiation of some of these cells into special organs. In particular they are believed to control development of some of the female sex characteristics. If the production of estrogens ever fails during pregnancy, the fetus dies almost immediately, and is expelled from the uterus.

Secretion of Progesterone during Pregnancy

Figure 347 also shows the amounts of progesterone secretion during pregnancy (expressed in terms of *pregnanediol,* which is the form in which progesterone is excreted in the urine). Progesterone, like estrogens, is secreted in moderate quantities by the corpus luteum during the early part of pregnancy. At approximately the sixteenth week of gestation the plancenta begins producing very large quantities of progesterone as well as estrogens, reaching a maximum production toward the end of pregnancy of as much as 10 times that produced by the corpus luteum.

Functions of progesterone during pregnancy. The first function of progesterone during pregnancy is to make increased quantities of nutrients available to the early endometrium for use by the developing ovum. It does this by causing the endometrial cells to store glycogen, fat, and amino acids. In addition, progesterone has a strong inhibitory effect on the uterine musculature, causing it to remain relaxed throughout pregnancy. It is believed that it is this effect that allows pregnancy to continue until the fetus is large enough to be born and live an independent existence. Despite these important functions of progesterone, some of the lower animals even among the mammals do not secrete this hormone.

Progesterone complements the effects of estrogens on the breasts. It causes the glandular elements to enlarge and to develop a secretory epithelium, and it promotes deposition of nutrients in the glandular cells, so that when milk production is required the appropriate materials will be available.

Progesterone affects the tubules of the kidney in much the same way that the adrenal mineralocorticoids affect them, to cause the reabsorption of sodium. Because of this function progesterone helps to increase the extracellular fluid volume and blood volume of the mother during pregnancy. However, the adrenal cortex also increases its production of mineralocorticoids during pregnancy, which adds to the increased fluids.

Function of Chorionic Gonadotropin during Pregnancy

If the corpus luteum is removed from the ovary at any time during the first three to four months of pregnancy, the loss of estrogen and progesterone secretion causes the fetus to stop developing and to be expelled within a few days. For this reason it is important that the corpus luteum remain active at least during the first third and preferably during the first half of pregnancy. Beyond that time removal of the corpus luteum does not affect pregnancy because the placenta by then is secreting many times as much estrogens and progesterone as the corpus luteum.

It will be recalled from the previous chapter that the corpus luteum normally degenerates and is absorbed at the end of each monthly sexual cycle. To keep the corpus luteum intact when the ovum implants, a special hormone, *chorionic gonadotropin,* is secreted by the developing fetal tissues—by the trophoblasts. This hormone has almost exactly the same properties as luteinizing and luteotropic hormones combined. It not only keeps the corpus luteum from involuting, but actually stimulates it so greatly that it enlarges progressively for the first 4 to 5 months of pregnancy.

Chorionic gonadotropin begins to be formed from the day that the trophoblasts implant in the uterine endometrium. Its concentration is highest approximately at the eighth week of pregnancy, as shown in Figure 347. Thus, the concentration is greatest at the time of pregnancy when it is essential to prevent involution of the corpus luteum. In the middle and later parts of pregnancy the secretion of gonadotropin falls to very low values. Its only known function at this time of pregnancy is to stimulate production of testosterone by the testes of the male fetus.

FETAL PHYSIOLOGY

In general, the physiology of the growing fetus is not far different from that of the normal child. All the organs of the body assume their final anatomic form, with minor exceptions, by four to five months after the beginning of pregnancy, and most of them can function almost completely normally by six months after the beginning of pregnancy. For example, long before birth the kidneys of the fetus become active, the gastrointestinal tract imbibes and absorbs fluid from the amniotic cavity, breathing is attempted even though it cannot be accomplished because the fetus is immersed in amniotic fluid, the heart pumps blood quite normally from approximately the third month on, and the metabolic systems operate very much the same as they do after birth.

Growth of the Fetus

Figure 348 depicts the increase in length and weight of the fetus during the 40 weeks of pregnancy. The length of the fetus increases almost directly in proportion to its age, while the weight increases in proportion to the third power of the age. The weight is almost infinitesimal until the twelfth to sixteenth week, but at that time it begins to increase extremely rapidly. Two months before birth the fetus usually weighs about one-half its birth weight, and one month prior to birth it weighs approximately three-fourths its birth weight. In other words, the major increase in weight occurs during the last two to three months. This is an especially important factor when one is considering the appropriate nutrition for the mother during pregnancy, because essentially all of the nutrients required by the baby are needed during the last three months. However, for several months prior to that time, growth of the uterus, placenta, and fetal membranes requires additional nutrients.

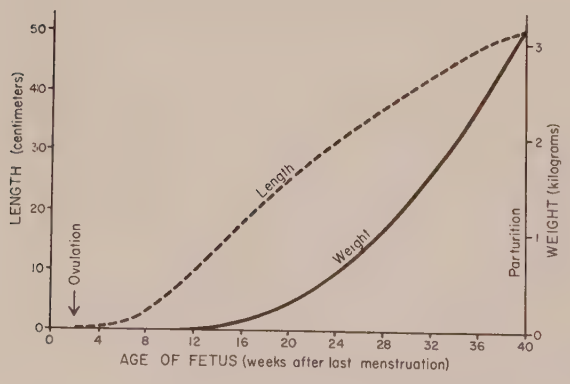


Figure 348. Growth of the fetus.

Special nutrients required by the fetus. The fetus requires especially large quantities of iron, calcium, phosphorus, amino acids, and vitamins. Iron begins to be used to form red blood cells within the first weeks of development of the fetus. Much of this early iron enters the fetus by active absorption from the endometrium by the trophoblasts. Throughout the remainder of pregnancy, still larger quantities of iron diffuse through the placental membrane to be used by the liver, spleen, and bone marrow for production of the fetus' blood.

Calcium is needed to ossify the bones. During the first two thirds of pregnancy the fetal bones contain mainly organic matrix and almost no calcium salts. During the last three months of fetal development, ossification occurs very rapidly, approximately one-half to two-thirds occurring in the final month. It is at this time, therefore, that the mother needs an especially abundant amount of milk or other foods containing calcium and phosphate.

The large quantities of amino acids and vitamins required by the fetus provide the necessary building materials for growth of the fetal tissues. During the last three months of pregnancy, a mother can become depleted of protein and vitamins if they are not in the diet in adequate amounts.

The Fetal Nervous System

The major anatomic parts of the nervous system are formed during the first few months of fetal growth, but complete function of this system is not reached even by the time the child is born. A premature baby always exhibits signs of poor nervous system function. As an example, one of the most difficult problems in treating a premature baby is to keep his body temperature normal, because the hypothalamic centers of a 6- or 7-month fetus usually have not developed sufficiently to regulate temperature. The child must be placed in an incubator for several weeks or sometimes a month or more until his nervous system has developed to a greater degree.

By the time of birth most, if not all, of the peripheral nerve fibers are completely developed, but in the the central nervous system the deposition of myelin around many of the large nerve fibers is far from complete. Though myelin is not always necessary for nerve fibers to function, it

has been inferred from this lack of complete myelination that certain portions of the central nervous system probably are far from functional at the time of birth. On the other hand, all the neurons that will ever be formed by the child are believed to be present by birth. This means that every time a neuron is destroyed thereafter, no new neuron takes its place.

Changes in the Circulation at Birth

Figure 349 illustrates the peculiarities of the circulatory system in the fetus. Special *umbilical vessels* allow blood to flow into and out of the placenta. The blood returning from the placenta through the umbilical veins enter the fetal circulatory system through a special vessel called the *ductus venosus,* which passes directly through the liver. Also, blood from the portal veins of the fetus flows through the ductus venosus, by-passing the liver entirely and emptying directly into the general venous system.

On reaching the heart, fetal blood by-passes the non-aerated lungs by two routes. First, it can flow from the right atrium through an opening called the *foramen ovale* into the left atrium, thus by-passing both the right ventricle and the lungs. Second, most of the blood that does not by-pass the right ventricle is pumped into the pulmonary artery and then through a vessel called the *ductus arteriosus* directly into the aorta.

Thus, by utilizing the ductus venosus, the foramen ovale, and the ductus arteriosus, blood flow through the liver and lungs occurs only to a minor extent in the fetus. This is a means for conserving the energy of the fetal heart, because the functions normally required of the liver will have been performed already by the mother's liver, and the fetal lungs are completely nonfunctional until birth.

After birth all three of these openings soon close. The ductus venosus and the ductus arteriosus become gradually occluded during the first few weeks of life. Presumably the ductus venosus becomes occluded because it no longer carries the tremendous blood flow from the umbilical veins, and the sluggish blood flow from the portal system is not sufficient to keep it open. The ductus arteriosus probably closes because of backward flow of blood from the aorta into the pulmonary artery. This in turn is caused by the following mechanism: As long as the lungs are collapsed in the fetus, very little blood flows through the lungs, and the principal flow is through the ductus arteriosus. However, when the lungs expand at birth, blood then begins to flow with ease through

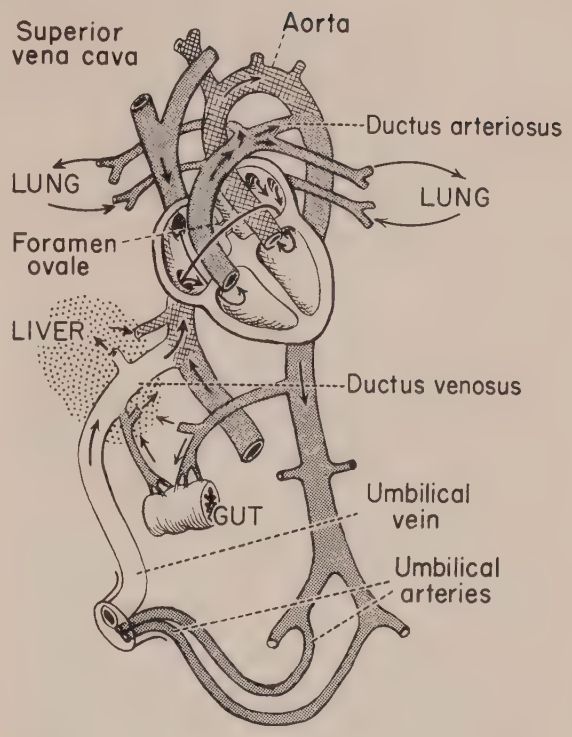


Figure 349. The fetal circulation, showing especially the ductus arteriosus, ductus venosus, and foramen ovale. (Modified from Arey: *Developmental Anatomy.* 6th Ed.)

them, causing the pulmonary arterial pressure to fall below the pressure in the aorta. Therefore, instead of flowing toward the aorta, the blood now flows in the opposite direction from the aorta into the pulmonary artery. For reasons not yet understood, this backward flow of blood causes the ductus to become occluded. Some research workers have claimed that valve-like folds along the wall of the ductus foster this occlusion.

The foramen ovale is covered by a thin valve on the left atrial side. As long as the pressure in the right atrium is greater than that in the left atrium, which is the case in the fetus, blood flows from right to left, by-passing the right ventricle and lungs. However, when the lungs expand, the right heart can then pump blood so easily through the lungs that the right atrial pressure falls to about 3 mm. Hg less than the left atrial pressure. This backward pressure differential threreafter keeps the valve of the foramen ovale closed. In about two thirds of all persons the valve gradually adheres to the opening so that it becomes permanently closed; in about one third it does not, but, regardless of which is true, blood ceases to flow through it after birth.

Thus, because of a series of changes in the dynamics of the circulation, blood begins to flow

through both the liver and lungs immediately after birth, even though in the fetus the heart had conserved its energy by by-passing these two organs.

PHYSIOLOGY OF THE MOTHER DURING PREGNANCY

The mother's physiology is changed during pregnancy in several ways. First, accessory changes occur in her reproductive organs and breasts to provide for development of the fetus and to provide nutrition for the newborn child. Second, all her metabolic functions are increased to supply sufficient nutrition to the growing fetus. Third, tremendous production of certain hormones by the placenta during pregnancy causes many side effects not directly associated with reproduction.

Changes in weight. The pregnant mother gains an average of about 20 pounds during pregnancy. In general, this gain is accounted for in the following manner: fetus, 7 pounds; uterus, 2 pounds; placenta and membranes, 2.5 pounds; breasts, 2 pounds; and the remainder is fat and increased quantities of extracellular fluid and blood. The amount of increase in these last two varies tremendously from one mother to another, depending upon her eating habits and upon the amounts of hormones secreted during pregnancy.

Changes in metabolism. The mother's metabolic rate in general increases approximately in proportion to her increase in weight, plus perhaps an additional 5 to 10 per cent. Much of this increase is occasioned simply by the increased amount of energy required for the mother to carry the growing load. However, rapid growth of the fetus also demands increased activity of most of the mother's functions, such as rapid intermediary metabolism in her liver, rapid pumping of blood by her heart, increased respiration, and increased digestion and assimilation of food.

Changes in the body fluids and circulation. The female sex hormones and the extra adrenocortical hormones produced during pregnancy cause the mother usually to gain about 5 to 7 pounds of fluid, or, in other words, about 3 liters. About 0.5 liter of this is in the plasma, and another 0.5 liter is red blood cells, making a total gain in blood volume of about 1 liter. About one third of the extra blood is needed to fill the sinuses of the placenta, but the other two thirds of a liter collects in the circulation, causing blood to flow toward the heart with far greater ease than usual. As a result, the cardiac output becomes roughly 25 per cent more than normal.

During birth of the baby, the mother loses an average of 200 to 300 ml. of blood as the placenta separates from the uterus. This ordinarily causes no physiological inconvenience because of the extra blood that has been stored during pregnancy.

BIRTH OF THE BABY (PARTURITION)

Duration of pregnancy and onset of parturition. The duration of pregnancy, between the time of the last menstrual period until birth of the baby, is normally 40 weeks, though occasionally surviving babies are born as early as 28 weeks or as late as 46 weeks. Approximately 90 per cent of all babies are born within 10 days before or after the 40-week interval.

The reason for the relatively constant duration of pregnancy has never been completely understood. Presumably, growth of the baby and of the placenta to a certain size and state of maturity initiates birth. The probable factors that cause the onset of parturition are the following:

When the baby becomes large, pressure of its body against the uterus stretches the uterine musculature, which in turn initiates uterine contractions. In addition to this, movements of the baby in the uterus—such as the feet and hands striking the uterine wall—also initiate contractions. Obviously, the larger the baby becomes, the more likely are these initiated contractions to become strong enough to cause birth.

Perhaps more important than the mechanical factors, however, are hormonal factors. The concentration of progesterone secreted by the placenta begins to decrease a few weeks prior to birth, and, since progesterone normally inhibits the uterus, this change allows an increase in uterine contractions. On the other hand, the concentration of active estrogens increases shortly before birth, and this increases the activity of the uterus, in contrast to the inhibition caused by progesterone. These two changes in hormonal secretion by the placenta probably account for much of the progressively increasing contractility of the uterus shortly before birth.

A third factor possibly helping to initiate parturition is the secretion of oxytocin by the hypothalamic-neurohypophyseal axis. This hormone causes extreme contractility of the uterine musculature, and absence of its secretion in animals usually makes parturition difficult.

Mechanism of parturition. Uterine musculature, like almost all smooth muscle, undergoes rhythmic contractions all the time. However,

because of the influence of progesterone, these contractions are so weak during the early months of pregnancy that they can hardly be noted. During the last three months of pregnancy, they increase steadily and reach extreme intensity a few hours before birth. These very strong contractions wedge the head against the cervix, stretch the cervical ring and vaginal canal, and expel the baby. This period, from the onset of the strong contractions until the baby is born, is called the period of *labor,* and the actual expulsion of the child is called *parturition.*

Approximately 19 times out of 20 the portion of the baby that pushes against the cervix is the head, and this acts as a wedge to open the cervical and vaginal canals. In most of the remaining cases the buttocks are the presenting portion of the baby, though occasionally a leg, a shoulder, or even the side of the baby may be against the cervical canal. In all of these instances the cervix cannot be wedged open nearly as effectively as by the head, and birth of the baby is considerably more difficult. When the head comes first, the remainder of the body slips through the vaginal canal within a few seconds after the head is born; when the head is not the presenting part, the body portion of the baby may be born relatively easily, but the head, the largest part of the child, then has difficulty in passing through the canal. One of the problems of bottom-first birth is often a period of interrupted umbilical blood supply to the baby, because the cord can become compressed against the wall of the delivery canal for a minute or more by the large head.

Theoretical cause of the sudden onset of labor. The reason the rhythmic uterine contractions suddenly become intense enough to cause labor is yet undetermined, but the following explanation has been offered as a possibility.

Irritation of the *cervix* is known to cause a muscular reaction over the entire uterus, making it become very excitable and causing its contractions to increase. It is presumed that when the natural contractions of the uterus reach a certain level of intensity, which occurs at the end of approximately nine months of pregnancy, they push the baby's head against the cervix, which stretches it and thereby makes the entire uterus contract more forcefully. This contraction in turn pushes the baby's head into the cervix still more and initiates still further intensity of uterine contraction. Thus, a vicious cycle develops, as illustrated in Figure 350, and the uterine contractions become stronger and stronger until finally the baby is expelled.

The increased uterine excitability caused by stretching or irritating the cervix is believed to result from two mechanisms. First, cervical stimulation is believed to transmit impulses upward through the uterine musculature itself, thus resulting in uterine contraction. Second, in lower animals and therefore presumably in man, cervical irritation transmits nerve impulses to the hypothalamus causing increased secretion of oxytocin, which in turn increases the contractility of the uterus, as was explained in Chapter 33.



Figure 350. A postulated mechanism for initiation of labor.

PRODUCTION OF MILK (LACTATION)

Hormonal Control of Lactation

The onset of secretion of the estrogens and progesterone after puberty in a girl begins to prepare the breasts for lactation. The breasts enlarge and the glandular elements begin to develop. However, the degree of glandular development is relatively slight compared with that achieved during pregnancy. The tremendous quantities of estrogens secreted by the placenta during pregnancy cause rapid development of the glands in the breasts, and the large quantities of progesterone change the glandular cells into actual secreting cells. By the time the baby is born, the breasts will have reached a degree of development capable of producing milk. Yet these same hormones, despite their developmental effects on the breasts, inhibit the actual formation of milk until after the baby is born.

Figure 351 shows the changes in hormone production at the time of parturition. Sudden expulsion of the placenta from the uterus removes

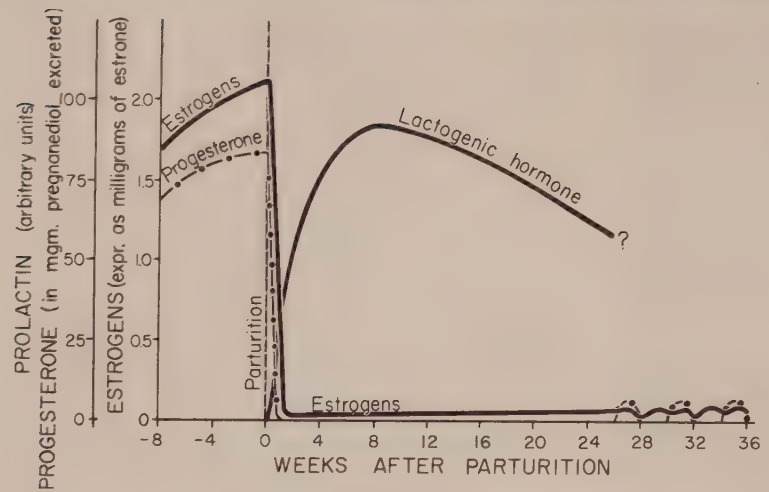


Figure 351. Changes in hormonal secretion that lead to the production of milk after parturition.

the source of estrogens and progesterone, so that the availability of these two hormones falls immediately almost to zero. Throughout pregnancy the production of gonadotropic hormones by the adenohypophysis has been inhibited by the placental hormones, but now that they are gone the adenohypophysis begins to secrete large quantities of *lactogenic hormone,* which is the same as *luteotropic hormone,* and which is also known as *prolactin.* This hormone causes the breasts to produce milk.

As long as the breasts are emptied of milk as rapidly as it is formed, the adenohypophysis will continue to produce lactogenic hormone, but when the milk is not emptied from the breasts, the production of lactogenic hormone also ceases. Though the reason for this is not known, it has been suggested that suckling sensations from the baby cause nerve impulses to pass from the nipples of the breasts, up the spinal cord, and then to the hypothalamus which secretes a *lactogenic hormone-releasing substance* into the hypothalamic-hypophyseal portal system to maintain lactogenic hormone formation.

As long as the adenohypophysis continues to produce lactogenic hormone it usually cannot produce sufficient follicle stimulation and luteinizing hormones to cause ovulation by the ovaries. Therefore, if the mother lactates, the normal female sexual cycle usually remains dormant for many months after parturition. But, if she does not lactate, her next ovulation appears about four weeks after birth and her next menstrual period about six weeks after birth.

Milk ejection. When a baby sucks on the nipples, it usually obtains no milk for approximately the first 45 seconds to a minute. Then suddenly milk appears in the ducts of both breasts even though suckling occurs on only one breast, indicating that some general phenomenon has occurred to cause milk to flow toward the nipples. Experiments have shown that suckling causes sensory impulses to pass first into the spinal cord, then upward through the brain stem, and finally into the hypothalamus to stimulate the production of oxytocin. This then circulates through the blood to the breasts, where it causes the myoepithelial cells surrounding the alveoli to contract and express the milk collected in the alveoli into the ducts leading to the nipples.

This milk ejection mechanism can be adversely affected tremendously by psychic factors. For example, a mother's fear that she might not be able to nurse her baby can actually keep her from doing so. Disturbances caused by other children in the family or by overly concerned relatives may lead to difficulty in milk ejection and cause the breasts to fail to empty. Failure to empty in turn allows the adenohypophysis to cease its production of lactogenic hormone, and the breasts stop secreting milk.

Composition of Milk

Milk contains the usual substances needed for energy and growth by the baby. These include two types of protein, *casein* and *lactalbumin;* an easily digested sugar, *lactose,* composed of one molecule of glucose and one of galactose; and large quantities of fatty substances such as *butter fats, cholesterol,* and *phospholipids.*

In addition, milk contains small quantities of vitamins and large quantities of calcium phosphate. Yet it has a conspicuous lack of iron. This lack usually is not detrimental to the baby in early life, because a sufficient quantity of this mineral is stored in the fetal liver prior to birth to continue

the formation of hemoglobin for about two months. Beyond this time, however, iron must be in the baby's diet, or he will develop progressive anemia.

Table 12 shows the relative compositions of human and cow's milk. When cow's milk is substituted for human milk, the baby's metabolic systems must adjust to a large increase in electrolytes—to the phosphates in particular. Also, it is desirable to fortify cow's milk with extra quantities of easily digested sugar such as pure glucose (dextrose).

Table 12. *Composition of Milk in Per Cent*

	Human	*Cow*
Water	88.5	87.0
Fat	3.3	3.5
Sugar	6.8	4.8
Casein	0.9	2.7
Lactalbumin and other protein	0.4	0.7
Electrolytes	0.2	0.7

Effect of Lactation on the Mother

Production of milk by the mother in some ways is as great a drain on her metabolic systems as pregnancy itself. Particularly does she lose tremendous amounts of stored proteins and fats during lactation. Also, if large quantities of calcium phosphate are not in her diet, her parathyroid glands become greatly overactive. This causes reabsorption of her bones, releasing the necessary calcium and phosphate for milk formation. As a result, her bones are likely to become weakened. Fortunately, though, the amount of calcium phosphate stored in the bones of the average mother is tremendous in comparison with the amount that the baby will need during the first few months of life. Therefore, only when the mother already has some degree of decalcification will this loss cause her any difficulty.

Frequently mothers develop extensive dental caries during pregnancy, which has often been ascribed to the loss of calcium phosphate from the teeth. Experiments, however, have shown that no significant amount of calcium phosphate leaves the teeth, but that the caries are probably caused by enhanced bacterial growth in the mouth rather than by any metabolic deficit.

REFERENCES

Arey, L. B.: *Developmental Anatomy.* 6th Ed. Philadelphia, W. B. Saunders Company, 1954.

Hagerman, D. D., and Villee, C. A.: Transport functions of the placenta. *Physiol. Rev.* 40:313, 1960.

Levi-Montalcini, R., and Angeletti, P. U.: Growth and differentiation. *Ann. Rev. Physiol.* 24:11, 1962.

Linzell, J. L.: Physiology of the mammary glands. *Physiol. Rev.* 39:534, 1959.

Nicholas, J. S.: Developmental physiology. *Ann. Rev. Physiol.* 22:95, 1960.

Parkes, A. S., editor: *Marshall's Physiology of Reproduction.* 3rd Ed. New York, Longmans, Green & Co., 1959.

Pincus, G.: Reproduction. *Ann. Rev. Physiol.* 24:57, 1962.

Potter, E. L.: *Fundamentals of Human Reproduction.* New York, McGraw-Hill Book Company, Inc., 1948.

Raven, C. P.: *Outline of Developmental Physiology.* New York, McGraw-Hill Book Company, Inc., 1954.

Saner, F. D.: *The Breast: Structure, Function, Disease.* Baltimore, The Williams & Wilkins Company, 1950.

Smith, C. A.: *Physiology of the Newborn Infant.* 3rd Ed. Springfield, Ill., Charles C Thomas, 1959.

Villee, C. A., and Villee, D. B.: Reproduction. *Ann. Rev. Physiol.* 25:263, 1963.

Walker, J., editor: *Oxygen Supply to the Human Foetus.* Oxford, Blackwell Scientific Publications, 1959.

INDEX

Absorption
  of fats, 338, 339
  of monosaccharides, 336
  of proteins, 342
Acceleration
  angular, 204
  in aviation, 203
  linear, 205, 207
Accidents, deceleratory forces and, 205
Acclimatization to oxygen deficiency, 202
Acetylcholine
  secretion at neuromyal junction, 223
  smooth muscle and, 235
Acid-base balance, 78
  abnormalities of, 81
  regulation of
    by kidneys, 80
    by respiration, 79
Acid-base buffers, 78
Acidosis, 81
  effect on body, 82
Acromegaly, 373
Actin, 227
Action potentials, 219
  conduction in heart, 121
Active immunity, 103
Active transport
  in renal tubules, 67, 68
  of potassium, 24
  through cell membrane, 23
  through placental villi, 410
Adaptation
  of eye, 301
  of sensory receptors, 248
  of smell, 317
Addison's disease, 382
Adenohypophyseal hormones, 372
Adenohypophysis, 372
Adenosine diphosphate, in control of energy metabolism, 36
Adenosine triphosphate, 347
  formation of, 25, 348
  in muscle contraction, 227
  properties of, special, 26
  uses in cell, 26
Adrenal cortex, 379
  sympathetic nervous system and, 284
Adrenal medulla, 285
Adrenergic nerves, 282
Adrenocortical hormones, 379
  amino acid flux and, 344
Adrenocortical secretion, abnormalities, 382
Agglutination of bacteria, 103
Agglutinins, 112
Agglutinogens, 112
Air, absorption from tissues, 199
Aldosterone
  effect on body, 379
  regulation by
    of arterial pressure, 161
    of sodium concentration, 75
  secretion of, regulation of, 380
Aldosterone stimulating substance, 381
Alkalosis, 81
  effect on body, 82
All or none law, 218
  in muscle contraction, 228
Allergens, 105
Allergic persons, 105
Allergic reactions, 105
Allergy, 100, 104
  hyperimmunization in, 106
Alveolar air, 183
Alveolar pressure, 180
Alveolar ventilation, 183
  in exercise, 194
  regulation of, 193
Ameboid motion, 29
  of neutrophils, 94
Amino acids
  absorption of, 342
  buffered by liver, 344
  energy from, 345
  for synthesis of substances, 345
  in blood, 343
  in protein synthesis, 28
  structural formulae of, 343
Ammonia excretion in acidic conditions, 80
Amygdaloid nuclei, emotions and, 291
Amylase, 335
Anaerobic metabolism, 25
Anaphylaxis, 105
Androgens, 382

Anemia, 92
  physiologic effects of, 93
Anemic anoxia, 195
Anesthesia, spinal, arterial pressure and, 158
Angiotensin, sodium concentration and, 76
Angular acceleration, 204
Anoxia, 195
  anoxic, 195
  histotoxic, 196
  oxygen therapy in, 197
Anterior motor neuron, 264
Antibodies, 101
  formation of, 102
  neutralization of, 103
Anticoagulants, extravascular, 110
Antidiuretic hormone, 374
  in control of osmotic pressure, 78
Antigen(s), 102
Antigen-antibody reactions, 102
Antiperistalsis, 327
Antithrombin, 109
Antithromboplastin, 109
Aortic regurgitation, pulse pressures in, 154
Aplastic anemia, 93
Appetite, 356
Aqueous humor, 62, 297
Arachnoidal granulations, 63
Arousal reaction, 288
Arterial pressure, 133
  carbon dioxide and, 159
  glomerular filtration and, 67
  in pulmonary circulation, 142
  mean, 156
  measurement of, 153
  medullary ischemia and, 159
  muscle blood flow and, 150
  pulsatile nature of, 152
  regulation of
    by kidney, 160
    during exercise, 161
    nervous, 157
  renin and, 160
  respiration and, 195
  systemic, 152
Arterial system, distensibility of, 154
Arteriolar dilatation by oxygen deficiency, 137
Arteriole(s)
  afferent, 66
  as regulators of blood flow, 137
  efferent, 67
  systemic pressures and, 156
Artificial kidney, 73
Artificial respiration, 180
Artificial respirator, 181
Ascorbic acid, 354
Assimilation of foods, 334
Asthma, 105, 198
Astigmatism, 299
Asynchronous summation, 231
Ataxia, 267
Atelectasis, 143, 198
Atherosclerosis, 146
  in diabetes, 386
Athetosis, 266
ATP. See *Adenosine triphosphate.*
Atrial fibrillation, 124
Atrial flutter, 124
Atrial pressure, right, 169
    venous return and, 166
Atrioventricular node, 121
Audiometer, 314
Auditory centers in brain, 313
Auditory cortex, 254, 313
Auscultatory method for pressure measurement, 153
Autologous tissue transplant, 115
Autonomic nervous system. See *Nervous system, autonomic.*
Autonomic reflexes of spinal cord, 275
Autoregulation, arteriolar, 137
A-V bundle, 121
  importance of, 122
Aviation physiology, 200
  decompression sickness in, 210
Axon reflex, 271

Bacteria, 13
Barometric pressure, effect on body, 200, 208
Basal ganglia
  abnormalities of, 266
  control of subconscious movements, 265
  function of, 265
Basal metabolic rate. See *Metabolic rate, basal.*
Bases, 78
Basilar fibers, 311
Basilar membrane, vibration of, 312
Basophils, 97
Bicarbonate buffer, 79
Bicarbonate ion, regulation in body fluids, 76
Bile
  secretion by liver, 332
  storage in gallbladder, 332
Birth of baby, 414
Blackout during positive g, 205
Bladder, urinary. See *Urinary bladder.*
Bleeding diseases, 110
Bleeding time, 111
Blind spot of eye, 304
Blood, 44
  cross-matching of, 111
  distribution in body, 139
  fatty acids in, 339
Blood cells, 87
  red. See *Red blood cells.*
  white. See *White blood cells.*
Blood clot, 108
Blood coagulation, 107
  mechanism of, 107
  tests for, 111
Blood flow, 132
  autoregulation in muscles, 150
  in lungs, 141
  in muscles, 150
  in systemic circulation, 131
  measurement of, 132
  nervous control of, 138
  regulation of, 137
  relation to pressure and resistance, 136
  resistance to, 135
  velocity of, 133
Blood groups, 112
Blood matching, 111
Blood pressure, 133
  high, 152
  in pulmonary circulation, 142
  measurement of, 133, 153
  relation to flow and resistance, 136
Blood reservoirs, 139
Blood substitutes, 114

Blood transfusion
  cardiac output and, 167
  mismatched blood in, 113
  techniques of, 113
Blood typing, 112
Blood vessels
  autonomic nerves and, 283
  clotting in, prevention of, 109
  diameter of, blood flow and, 135
  endothelium of, repulsion of platelets by, 109
  resistance of, venous return and, 166
  spinal cord control of, 275
  vascular spasm of, in hemostasis, 109
Blood viscosity, blood flow and, 135
Blood volume
  calculation of, 51
  regulation of, 82
Blue baby, 145
Blue weakness, 303
Body fluid(s), 43
  constituents of, regulation of, 75
  extracellular, 43
    constituents of, 44
    mixing of, 44, 47
    osmotic equilibrium with intracellular fluid, 49
  in pregnancy, 414
  interstitial, 44
    colloid osmotic pressure of, 54
      in edema, 60
  intracellular, 43
    constituents of, 44
    osmotic equilibrium with extracellular fluid, 49
  regulation of, 75
  solutions for injection into, 114
  special systems, 60
  volume(s)
    extracellular, measurement of, 50
      regulation of, 82
    interstitial, calculation of, 51
    intracellular, calculation of, 51
      regulation of, 83
Body temperature, 358, 363
  blood flow and, 150
  regulation of, 364
Body water, measurement of, total, 50
Bone, 387, 389
  chemical composition of, 389
  deposition of, 390
  reabsorption of, 390
    parathyroid hormone in, 391
Brain. See also *Cortex, cerebral.*
  auditory centers in, 313
  connections with retina, 303
  control by
    of eye movements, 306
    of motor functions, 256
  dominance in, hemispheric, 255
  gnostic area of, 255
  in headache, 254
  integrative area of, common, 255
    control of motor functions, 256, 262
  interpretation of sensations by, 246, 254
  muscle control by, 257
  smell sensations to, transmission of, 318
  taste sensations to, transmission of, 316
Brain stem, 270
  control of autonomic nervous system by, 286
  motor functions of, 276
Brain waves, 289
  abnormal patterns of, 290
Breathing, regulation of, 191
Broca's area, 269
Bulboreticular formation, 276
Bursae, 60

Calcium, in muscle contraction, 226
Calcium ions, function of, 387
Calcium metabolism, 387
  regulation by parathyroid hormone, 391
Calorie, 350
Calorimeter for humans, 360
Cancer, 37
Capillaries
  anatomy of, 52
  diffusion through, 47
  in arterial pressure regulation, 160
  in blood volume regulation, 82
  law of, 55
  lymphatic, 56
  pressures in, 53
    in edema, 59
  pulmonary, 143
    pressure in, 142
  transit time for blood in, 133
Capillary membrane
  dynamics of, 52
  equilibrium at, 55
  pressure gradient across, 53
Capillary porosity, in edema, 60
Carbaminohemoglobin, 189
Carbohydrates, 334
  digestion of, 335
  fat-sparing effect of, 341
  formed from protein and fat, 346
  in ATP formation, 348
  in protoplasm, 15
  oxidative release of energy by, 348
  protein-sparing effect of, 346
Carbon dioxide
  alveolar ventilation and, 193
  arterial pressure and, 159
  chemical combinations with blood, 189
  deep sea diving and, 209
  diffusion in tissues, 186
  high altitudes and, 202
  pressure of, partial, in alveoli, 184
  respiratory regulation by, 193
  transport of, 177, 189
    through pulmonary membrane, 185
Carbon monoxide, as hemoglobin poison, 188
Carbonic anhydrase, 189
  in transport of carbon dioxide, 189
Cardiac constrictor, in swallowing, 324
Cardiac cycle, 127
Cardiac failure, 164, 171
Cardiac impulse
  block of, 122
  conduction in heart, 121
  delay at A-V node, 122
Cardiac muscle, 232
  rhythmicity of, 232
Cardiac output, 164
  blood transfusion and, 167
  equilibrium with venous return, 167
  in heart failure, 171
  measurement of, 168
  regulation of, 165
    summary of, 168
  values of, 164

Cardiac pressures during heart cycle, 127
Cardiac rhythmicity, 120
Cardiac shock, 172
Caries, 395
Catalysis of cellular substances, 29
Caudate nucleus, 265
Cell(s)
  ameboid motion of, 29
  blood, 87. See also *Red blood cells* and *White blood cells.*
  control of function, 32
  differentiation of, in embryo, 39
  energy metabolism control in, 36
  glycogen in, 19
  hair
    in cochlea, 311
    in utricle, 277
  internuncial, 264
  Kupffer, in liver, 100, 148
  myoepithelial, of breast, 375
  organelles of, 17
  organization of, general, 16
  oxidation in, ADP and, 349
  physical structure of, 15
  reproduction of, 32, 37
  respiration of, 25
  reticulum, 100
  synthesis in, 26
  types of, 12
Cell growth
  control of, 36
  inhibition of, 37
Cell membrane, 16
  osmosis at, 48
  pores in, 16
  transport through, 21
    active, 23
Cell reproduction, 37
Cellular respiration, 25
Cementum of teeth, 393
Central nervous system. See *Nervous system, central* and *Brain.*
Centrioles, 19
Cerebellum
  function of, 266
    damping, 267
    equilibrium, 268, 278
    predictive, 268
Cerebral circulation, 146
  regulation of, 147
Cerebral cortex. See *Cortex, cerebral* and *Brain.*
Cerebral vascular accidents, 148
Cerebrospinal fluid, formation of, 62
Cerebrospinal fluid system, 62
  blockage of, 63
Chemical senses, 315
Chemoreceptor system, 194
Chemotaxis, 29
  of neutrophils, 95
Chills, 356
Chloride ion
  aldosterone and, 380
  regulation in body fluids, 75
Cholesterol, 341
Cholinergic nerves, 282
Cholinesterase, destruction of acetylcholine by, 223
Chorda-mesoderm, 40
Chorea, 266
Chorionic gonadotropin, 411
Choroid plexuses, 62
Chromosomes
  genes in, 33
Chromosomes (*Continued*)
  in ovum formation, 402
  in sperm formation, 398
  replication of, 38
Chylomicrons, 339
Chyme, 325
  neutralization of, by pancreatic secretion, 330
Cilia, 30
Ciliary motion, 30
Ciliary muscle of eye, 307
Circulation. See also *Heart* and *Circulatory system.*
  angular acceleration and, 204
  cerebral, 146
  coronary, 145
  in placenta, 409
  pulmonary, 141
Circulatory shock. See *Shock, circulatory.*
Circulatory system. See also *Heart* and *Blood vessels.*
  aldosterone and, 380
  fetal, changes at birth, 413
  in pregnancy, 414
  integration in body, 6
  portal, 148
  schema of, general, 47
  transport of extracellular fluid, 45
Circus movement in heart, 123
Clasmatocytes, 101
Clothing, heat loss and, 363
Clotting time, 111
Cobalt, 356
Cochlea, 310
  anatomy of, 310
  nerve impulses from, 313
  resonance of sound in, 311
Coenzyme, 348
Colloid osmotic pressure, 54
  at capillary membrane, 54
  in edema, 59
  of interstitial fluid, regulation by lymphatics, 57
  of plasma, 54
Colloids in protoplasm, 20
Colon, movements of, 326
Color blindness, 303
Color vision, 302
  intermediate colors and, 303
Common integrative area of brain, 255
  control of motor functions, 256, 262
Communication, interpretation by brain, 269
Conduction, for heat loss, 362
Cones, of retina, 300
    color vision and, 302
    visual acuity and, 305
Congenital heart disease, 143
Consciousness, thalamus in, 249
Copper, 356
Cornea, 297
  transplantation of, 115
Coronary circulation, 145
  regulation of, 145
Coronary occlusion, 146
Cortex, cerebral
    control of autonomic nerves, 286
    control of speech, 268
    emotions and, 290
    thalamus and, 256
Corticospinal tract, 264
Corticotropin, 373, 382
Corticotropin releasing factor, 382
Cough, 179
Creatinine, lack of reabsorption in tubules, 69

Crystalloidal osmotic pressure, 77
Curare, at neuromyal junction, 224
Cytochrome, 349
Cytoplasm, 17

Dark adaptation, 301
Dead space, 183
Deafness, 314
Deceleration, 205
  of spaceship, 207
Decerebrate rigidity, 277
Decompression, explosive, 203
Decompression sickness, 210
  prevention of, 210
Deep sea diving physiology, 200, 208
Defecation, 326
Dehydration, 50
Dental abnormalities, 395
Dentine, 393
Dentition, 394
Deoxyribonucleic acid(s), 19
  in protein synthesis, 27, 33
  repression of, 35
Depolarization wave, 217
Depth perception, 306
Dermatomes, 250
Desensitization to allergens, 106
Dextran, as blood substitute, 114
Diabetes, 385
  atherosclerosis in, 386
  pituitary, 383
Diabetic coma, 385
Diastole, 127
Diastolic pressure, 152
Diet, regulation by taste, 316
Diffusion
  facilitated, 23
  in renal tubules, 68
  of fluids and solutes, 46
  of gases
    in tissues, 185
    in water, 185
  principle of, 46
  through membrane(s), 21
    of placenta, 409
    pulmonary, 186
Diffusion gradient, 23, 47
Digestion, 334
  by neutrophils, 96
  of carbohydrates, 335
  of fats, 338
  of proteins, 342
Dilution principle for fluid volumes, 50
Diuretics, kidney output and, 72
Diving physiology, 208
DNA. See *Deoxyribonucleic acid(s).*
Drinking center, 83
Ductus arteriosus, 413
Ductus venosus, 413
Duodenocolic reflex, 327
Dwarfism, 373
Dye method for cardiac output measurement, 168
Dyspnea, 197

Ear, inner, 309
Edema, 59
  in heart failure, 171
  in potential spaces, 61
  pulmonary, 127, 198
Ejaculation, 399
Electrocardiogram, 124
Electroencephalogram, 289
Electrolytes
  absorption of, 346
  appetite for, 83
  in protoplasm, 14
  membrane potentials and, 216
  regulation in body fluids, 75
Emotions, 290
Emphysema, 198
Enamel of teeth, 393
Endemic goiter, 378
Endocrinology, 371
Endometrial cycle, 405
Endoplasmic reticulum, 17
Endplate potential, 226
Energy
  daily requirements, 351
  for muscle contraction, 227
  from amino acids, 345
  from fats, 340
  from foods, 347, 350, 351
    schema for transfer of, 350
  from glucose, 338
Energy metabolism
  of cell, 25
    control of, 36
Enzyme(s)
  control of, 35
  hydrolytic, 29
  in protoplasm, 15
  thyroxine and, 376
Enzyme inhibition, 35
Eosinophils, 97
Epididymis, 399
Epilepsy, 290
Epinephrine, blood glucose and, 337
Equilibrium
  brain stem and, 277
  cerebellum and, 268
Erection
  in female, 403
  in male, 399
Erythroblastosis fetalis, 113
Erythropoietic stimulating factor, 92
Erythropoietin, 92
Esophagus
  in swallowing, 324
  secretion in, 329
Essential hypertension, 163
Estrogens, 404
  in pregnancy, 410
Evaporation, for heat loss, 362
Excitatory neurons, 240
Excitatory postsynaptic potential, 238
Excitatory transmitter, 238
Excretion through placenta, 410
Excretory system, function of, 9
Exercise
  alveolar ventilation in, 194
  arterial pressure in, regulation of, 161
  cardiac output in, 167
  muscle hypertrophy in, 232
Exophthalmos, 377
Expiratory reserve volume, 182
Extensor thrust reflex, 273
Exteroceptive receptors, 248
Extracellular fluid. See *Body fluid(s), extracellular.*
Eye, 297

Eye (*Continued*)
autonomic nerves and, 282
blind spot of, 304
ciliary muscle of, 307
depth perception by, 306
fluid systems of, 61
focusing of, 307
lens system of, 298
optics of, visual acuity and, 305
pressure in, 62
pupil of, control of, 308
retina of, 300
Eye movements, 306
moving objects and, 307

Facilitated diffusion, 23
Facilitation at synapse, 239
Fallopian tubes, 402
Fat(s), 334
conversion to carbohydrates, 346
coronary disease and, 146
digestion of, 338
energy from, 340
hydrocortisone and, 381
in ATP synthesis, 349
insulin and, 384
synthesis of, 340
transport in body fluids, 339
unsaturated in diet, 351
Fat tissue, 339
Fatty acids, 338
in blood, 339
Feedback, in control of cell function, 32
Ferritin, 91
Fetal physiology, 407, 412
Fetus
circulatory changes at birth, 413
growth of, 412
nervous system of, 412
nutrition of, 408, 412
Fever, 356
value of, 367
Fibrillation
atrial, 124
ventricular, 124
Fibrin, 108
Fibroblasts, conversion from lymphocytes, 98
Fick principle, 168
Filtration pressure, 56
in glomerulus, 65
Flavoprotein, 348
Flexor reflex, 273
Fluid(s). See *Body fluid(s).*
Fluid flow in tissues, 55
Fluid systems of eye, 61
Fluid volumes. See also *Body fluid(s), volume(s).*
dilution principle for, 50
measurement of, 50
Fluorine, 356
Flutter, atrial, 124
Focusing of eyes, 307
Folic acid, 354
in red cell formation, 90
Follicle stimulating hormone, 374, 400, 403
Follicles of ovary, 401
Food
assimilation of, 334
intake of, regulation of, 356
specific dynamic action of, 359
vitamin contents of, 353
Foramen ovale, 413
Functional residual capacity, 182

g forces, 204
Gallbladder
emptying of, 332
storage of bile in, 332
Gallstones, 332
Gamma aminobutyric acid, 239
Gas(es)
diffusion of
in tissues, 185
in water, 185
Gastric. See *Stomach.*
Gastrin, 330
Gastrocolic reflex, 327
Gastrointestinal movements, 320, 322, 327
Gastrointestinal secretions, 328
Gastrointestinal system
autonomic nerves and, 283
integration in body, 8
Gastrointestinal tract, 320
mixing movements in, 323
peristalsis in, 323
thyroxine and, 376
Genes, 19, 33
protein synthesis control by, 33, 344
replication of, 38
Giantism, 373
Glands, transmission of impulses, 223
Glaucoma, 62
Globus pallidus, 265
Glomerular filtrate, 67
Glomerular filtration, 65
Glomerular membrane, fluid dynamics at, 65
Glomerular pressure, 65
Glomerulonephritis, 72
Glomerulotropic hormone, 381
Glomerulotropin, in regulation of sodium concentration, 76
Glomerulus, 64
Glucagon, 386
Glucocorticoids, 381
Gluconeogenesis, 337
Glucose
distribution in body, 335
fate in tissues, 336
hydrocortisone and, 381
in fat synthesis, 340
insulin and, 384
insulin production and, 337
Glucose tolerance test, 384
Glycogen in cells, 19
Gnostic area, 255
Goiter, 378
Golgi complex, 18
Gonadotropic hormones, 373, 403
Gravity, support of body against, 276
Growth hormone, 372

Hair cells
in cochlea, 311
in utricle, 277
Hay fever, 105
Headache, 253
Hearing, 309
Hearing tests, 314

Heart. See also *Cardiac; Circulation;* and *Circulatory system.*
autonomic nerves and, 283
circus movement in, 123
cycle of, 127
law of, 128
musculature of. See *Heart muscle.*
nervous control of, 129
pumping action of, 119
effectiveness of, 165
refractory period of, 122
regulation of, 119
rhythmicity of, 120, 233
vascular supply of, 145
Heart disease, congenital, 143
Heart failure
cardiac output in, 171
edema in, 59
Heart-lung preparation, 129
Heart muscle
impulse transmission in, 222
rhythmicity in, 120, 233
Heart pressures during cardiac cycle, 127
Heart sounds, 125
Heart valves
diseases of, 126
function of, 125
regurgitation at, 127
stenosis of, 126
Heat
liberated by muscle, 228
loss from body, 362
production in body, 358
loci for, 361
Heat reflexes, 285
Hematocrit, 87
Hemiballismus, 266
Hemodynamics, 131
Hemoglobin
as oxygen buffer, 188
carbon dioxide and, 189
fate of, 91
formation of, 90
of red blood cell, 88
oxygen transport by, 187
Hemoglobin reserve capacity, 188
Hemophilia, 110
Hemopoietin, 92
Hemorrhage
anemia from, 92
shock from, 172
Hemostasis, 109
Heparin, 109
Hering-Breuer reflex, 192
Heterologous transplant, 115
High altitudes, oxygen deficiency and, 201
High blood pressure, 152
High energy bonds, 26
High pressure physiology, 208
Histiocytes, 100
Histotoxic anoxia, 196
Homeostasis, 4
Homologous transplant, 115
Hormonal hypertension, 162
Hormone(s), 9, 371
adenohypophyseal, 372
adrenocortical, 379
antidiuretic, 374
female, 403
follicle stimulating, 374, 400, 403
general, 371
glomerulotropic, 381
Hormone(s) (*Continued*)
gonadotropic, 373, 403
growth, 372
hypophyseal, 372
in control of lactation, 415
lactogenic, 416
local, 371
luteinizing, 374, 400, 403
luteotropic, 374, 403, 416
neurohypophyseal, 374
of pregnancy, 410
ovarian, 404
parathyroid, 387, 391
sex
female, 403
male, 400
thyroid. See *Thyroid hormone.*
Humidification of inspired air, 184
Humoral immunity, 103
Hunger, 356
Hunger contractions, 325
Hydrocortisone, 381
regulation of secretion, 382
Hydrogen ion concentration, control of respiration, 79, 193
Hydrolytic enzymes, 29
Hydrostatic pressure, venous pressure and, 169
Hypercalcemia, 392
Hyperimmunization in allergy, 106
Hyperinsulinism, 386
Hypermetropia, 299
Hypertension, 152, 162
essential, 163
neurogenic, 163
renal, 162
Hyperthermia, 367
Hyperthyroidism, 377
Hypertonic solution, 49, 50
Hypocalcemia, 392
Hypophyseal hormones, 372
Hypophysis, 371
Hypothalamic-hypophyseal portal system, 374
Hypothalamus
control of autonomic nerves by, 286
emotions and, 291
in body temperature regulation, 364, 365
proteins and, 366
Hypothermia, 367
Hypothyroidism, 378
Hypotonic solution, 49, 50
Hypovolemic shock, 172

Ideomotor area, 259
dominance of, 260
Ileocecal valve, 326
Immune bodies, 102
Immune process, 102
Immune reaction, 104
allergy and, 104
Immune response, 104
Immunity, 100, 101
humoral, 103
passive, 103
Impulse transmission at synapse, 238
Induction, 40
Inflammation, 101
Inhibition at synapse, 239
Inhibitory neurons, 240
Inner ear, 309
Inspiratory capacity, 182

Insulin, 379, 383
  blood glucose and, 384
  fat metabolism and, 384
  glucose transport and, 336
  in diabetes, 386
  mechanism of action, 383
  proteins and, 384
  regulation of secretion, 384
Internuncial cells, 264
Interstitial fluid. See *Body fluid(s), interstitial.*
Intestine
  large, secretions of, 333
  small
    absorption from, 332
    movements of, 325
      propulsive, 326
  secretions of, 332
Intracellular fluid. See *Body fluid(s), intracellular.*
Intramural plexus of gut, 322
Intrapleural pressure, 180
Intrapleural space, 60
Iodine, 355
Ion
  bicarbonate, regulation in body fluids, 76
  chloride
    aldosterone and, 380
    regulation in body fluids, 75
  diffusion through membranes, 22
Iron, 355
  deficiency of, anemia in, 92
  in formation of hemoglobin, 90
Iron metabolism, 90
Irreversible shock, 174
Ischemia, medullary, arterial pressure and, 159
Isotonic solution, 50
  infusion of, 50

Joint spaces, 60

Ketosis, 385
Kidney. See also *Renal.*
  abnormal function of, 71
  anatomy of, 64
  antidiuretic hormone control of, 374
  arterial pressure regulation by, 161
  artificial, 73
  as cause of hypertension, 162
  blood volume regulation by, 82
  clearance by, 70
  counter-current mechanism of, 77
  loss of nephrons, 71
  output of, diuretics and, 72
  regulation of acid-base balance by, 80
  shutdown of, 71
  sympathetic stimulation of, 284
Kidney function, renal clearance as test of, 70
Knee jerk, 272
Knowledge, 259
Kupffer cells of liver, 100, 148

Labor, childbirth, 415
Lactation, 415
  effect on mother, 417
Lactogenic hormone, 416
Larynx
  function of, 178
  testosterone and, 401
Law(s)
  of capillaries, 55
  of gut, 323
  of heart, 128
  of specific nerve energies, 247
  Weber-Fechner, 251
Learning, reward and punishment and, 291
Lens, convex, 298
Lens system of eye, 298
Leukemia, 98
Leukopenia, 99
Leukotaxine
  as cause of chemotaxis, 95
  in inflammation, 101
Lever systems of body, 230
Life, precellular, 13
Light adaptation, 301
Linear acceleration, 205
  of spaceship, 207
Lipids. See also *Fat(s).*
  diffusion through membrane, 22
  in protoplasm, 15
Lipoproteins, 340
Liver
  as blood reservoir, 139
  buffer effect on glucose, 337
  circulation of, 148
  in fat utilization, 340
  in plasma protein formation, 344
  integration in body, 9
  Kupffer cells of, 100, 148
  release of glucose, 284
  secretion by, 331
Locomotion, control of, 279
Loop of Henle, urine concentration and, 77
Lungs
  autonomic nerves and, 283
  blood flow in, 141
  dryness of, maintenance of, 143
  in respiration, 177
  respiration in, 179
Luteinizing hormone, 374, 400, 403
Luteotropic hormone, 374, 403, 416
Lymph, 56
  filtration by lymph nodes, 58
  rate of flow, 58
Lymph flow, 57
  from potential spaces, 61
Lymph nodes, filtration of lymph by, 58
Lymphatic capillaries, 56
Lymphatic system, 56
  fat transport through, 339
  function of, 57
  integration in body, 7
Lymphocytes, 97
Lymphocytic leukemia, 99
Lysis of bacteria, 103
Lysosomes, 18
  in catalysis of cellular substances, 29

Magnet reaction, 273
Malocclusion, 395
Manometer, mercury, 133
Margination of neutrophils, 96
Maximal stimulus of nerve, 219

Mean arterial pressure, 156
  regulation of, 156
Mean systemic pressure, venous return and, 166
Medullary ischemia, arterial pressure and, 159
Meissner's plexus of gut, 322
Membrane
  cell. See *Cell membrane.*
  nerve
    depolarization of, 216
    electric potentials of, 215, 216
    repolarization of, 217
  nuclear, 20
  pulmonary, 185
  tympanic, 310
Memories, 258
  synapses and, 240
Memory process, 258
Menstruation, 405
Mercury manometer, 133
Mesencephalon, in motor control, 262
Metabolic rate, 358
  basal, 359
    determining, direct method for, 360
  thyroxine and, 376
Metabolic systems, integration in body, 8
Metabolism
  anaerobic, 25
  autonomic nerves and, 284
  in pregnancy, 414
  iron, 90
Microsomes, 18
Micturition, 64, 73
Micturition reflex, 74
Milieu interne, 4
Milk
  composition of, 416
  ejection from breasts, 375, 416
  production of, 415
Mineralocorticoids, 379
Minerals of body, 355
Minimal stimulus of nerve, 219
Minute respiratory volume, 182
Mitochondria, 17
  structure of, 18
Mitosis, 37
  factors that determine, 38
  rates of, 38
  regulation of, 38
  stages of, 38
Modality of sensation, 249
Monocytes, 97
Monophasic action potential, 220
Monosaccharides
  absorption of, 336
  fate in tissues, 336
Motor control, 257
Motor cortex, 259, 261
Motor functions
  of brain stem, 276
  overall control of, 262
Motor movements, coordination by cerebellum, 266
Motor unit, 228
Motor unit summation, multiple, 231
Mucosal reflexes, 327
Mucus, 329
Multiple factor theory for respiration, 195
Muscle
  atrophy following denervation, 232
  blood flow in, 150
    sympathetic nerves and, 285
Muscle (*Continued*)
  cardiac, rhythmicity of, 120, 233
  contractile process of, 227
  contraction, summation of, 231
  control by brain, 257
  heart. See *Heart muscle.*
  hypertrophy in exercise, 232
  impulses in, transmission of, 222
  integration in body, 4
  of respiration, 180
  signals to, transmission of, 263
  skeletal, 225, 232
  smooth, 232
    contractile process in, 234
    control of contraction, 235
    plasticity of, 234
    rhythmicity in, 234
Muscle contraction
  control by cortex, 259, 262
  coordination of antagonistic muscles, 277
  duration of, 229
  efficiency of, 228
  initial length and, 230
  isometric, 229
  isotonic, 229
Muscle function, overall control of locomotion, 279
Muscle physiology, 225
Muscle pump, 169
Muscle reflexes, 272
Muscle tone, stretch reflex and, 273
Muscle twitch, 229
Myasthenia gravis, 224
Myelin sheath, 221
Myelinated fibers, 221
Myelogenous leukemia, 99
Myenteric plexus of gut, 322
Myoepithelial cells of breast, 375
Myofibrils, 226
Myoneural junction, 223
Myopia, 299
Myosin, 227

Neck proprioceptors, in equilibrium, 278
Necrosin, in inflammation, 101
Negative g, 204
Neostigmine, at neuromyal junction, 224
Nephron(s), 64. See also *Kidney.*
  function of, 65
    summary of, 69
  loss of, 71
Nerve(s)
  adrenergic, 282
  cholinergic, 282
  function of, 215
  maximal stimulus of, 219
  minimal stimulus of, 219
  parasympathetic
    control of heart, 129
    to gut, 322
  sympathetic
    control of heart, 129
    to gut, 322
Nerve conduction, velocity of, 221
Nerve endings, free, 248
Nerve energies, specific, law of, 247
Nerve excitation, all or none law and, 218
Nerve fibers
  anatomy of, 215

Nerve fibers (*Continued*)
in nerve bundles, 221
law of specific nerve energy, 247
refractory period of, 218
Nerve impulse, 216, 217
recovery from, 218
transmission of, 217
over nerve trunk, 221
Nervous control
of food intake, 356
of swallowing, 323
Nervous regulation of arterial pressure, 157
Nervous system. See also *Brain.*
autonomic, 6, 280
control by brain, 286
control of smooth muscle, 235
integration in body, 6
organs and, 283
blood flow control by, 138
central
salivation and, 328
smell sensations into, transmission of, 318
sound transmission into, 313
stomach secretion and, 329
heart and, 129, 165
integration in body, 5
of fetus, 412
organization of, basic, 236
parasympathetic, 280
anatomy of, 282
control of smooth muscle, 235
sympathetic, 280
anatomy of, 281
control by
of muscle blood flow, 150
of smooth muscle, 235
function of, overall, 284
metabolism and, 358
thyroxine and, 376
Neurogenic hypertension, 163
Neurohypophyseal hormones, 374
Neuromyal junction, 223
Neuron(s)
discharge of, repetitive, 239
inhibitory, 240
integrative function of, 241
of sympathetic nervous system, 281
postganglionic, 281
preganglionic, 281
response characteristics of, 239
synaptic functions of, 236
Neuronal circuits, 241
converging, 242
diverging, 242
integrative function of, 243
parallel, 243
Neuronal pool, 241
facilitated zone in, 241
input versus output, 244
Neutralization of antibodies, 103
Neutrophilia, 96
Neutrophils, 94
chemotaxis of, 95
phagocytosis by, 96
Niacin, 354
Nitrogen, in decompression sickness, 210
Nitrogen poisoning, 209
Nociceptive reflex, 273
Node of Ranvier, 221
Norepinephrine, smooth muscle and, 235
Nose, function of, 178
Nuclear membrane, 20
Nucleic acid, autocatalytic nature of, 14
Nucleolus, 19
Nucleoproteins in protoplasm, 15
Nucleus, 19
Nutrition, 347, 350
of fetus, 408, 412

Obesity, 357
Oculomotor centers, 306
Odors, masking of, 317
Olfactory receptors, 317
Optic disc, 304
Organelles of cells, 17
Osmoreceptors, in control of osmotic pressure, 78
Osmosis
at cell membrane, 48
in renal tubules, 68
principle of, 48
Osmotic pressure, principle of, 54
Ossicles, 310
Osteoclasts, 390
Ovarian hormones, 404
Ovulation, 406
Ovum, 401
early division of, 408
fertilization of, 407
implantation of, 408
transport of, 402
Oxygen
absorption from tissues, 199
diffusion in tissue spaces, 186
hemoglobin buffering of, 188
in regulation of coronary blood flow, 145
lack of, red cell production and, 91
partial pressure in alveoli, 184
space trips and, 207
transport of, 177
by red blood cells, 88
in tissues, 187
through pulmonary membrane, 185
Oxygen breathing, at high altitudes, 203
Oxygen debt, 350
Oxygen deficiency
acclimatization to, 202
at high altitudes, 200
autoregulation of blood flow in, 137
effects on body, 201
in regulation of respiration, 194
Oxygen poisoning, 209
Oxygen therapy, 197
Oxytocin, 374

Pain, 249
perception of, 249
psychosomatic, 293
reactivity to, 249
referred, 252
visceral, 252
Pain reflex, 273
Pancreas
insulin secretion by, 383
regulation by vagus, 331
Pancreatic secretion, 330
Pancreozymin, 331
Panting, 365
Pantothenic acid, 354
Parachute descent, 205

Parasympathetic nerves. See *Nerves, parasympathetic.*
Parasympathetic nervous system. *See Nervous system, parasympathetic.*
Parasympathetic tone, 285
Parathyroid hormone, 387, 391
Parathyroid secretion
  abnormalities of, 392
  regulation of, 392
Parkinson's disease, 266
Partial pressure of respiratory gases, 183
Partial proteins, 351
Parturition, 414
  mechanism of, 414
Passive immunity, 103
Patent ductus arteriosus, 143
  pulse pressure in, 154
Pericardial space, 60
Peripheral congestion in heart failure, 171
Peripheral resistance
  venous return and, 166
  total, 157
Peripheral venous pressure, 169
Peristalsis, 323
Peritoneal reflex, 275, 327
Peritoneal space, 60
Pernicious anemia, 93
Phagocytosis
  by neutrophils, 96
  in spleen, 149
  of bacteria, 103
Phonocardiogram, 126
Phosphate buffer, 79
Phosphate ions
  relation to calicum ions, 388
  relation to parathyroid hormone, 391
Phosphocreatine, 349
Phospholipids, 341
Physostigmine, at neuromyal junction, 224
Pilo-erection, 365
Pinocytosis, 24
Pituitary diabetes, 383
Placenta, 408
Plasma
  as blood substitute, 114
  colloid osmotic pressure of, 54
Plasma cells, conversion from lymphocytes, 98
Plasma protein, formation of, 344
Plasma volume, measurement of, 51
Platelet(s), 108
Platelet count, 111
Plethysmograph, 132
Pneumonia, 197
Pneumotaxic center, 192
Pneumothorax, 180
Poiseuille's law, 136
Poisoning
  nitrogen, 209
  oxygen, 209
Polycythemia, 93
  physiologic effects of, 94
Pores
  diffusion through, 22
  in cell membrane, 16
Portal circulatory system, 148
Portal obstruction, 149
Positive after potential, 220
Positive g, 204
Postganglionic neurons, 281
Potassium
  active transport of, 24
  aldosterone and, 380
Potassium (*Continued*)
  regulation in body fluids, 76
Potential fluid spaces, 60
  lymphatic drainage from, 61
Prefrontal area, function of, 260
Prefrontal lobes, loss of, 261
Preganglionic neurons, 281
Pregnancy, 407
  hormones of, 410
  metabolism in, 414
  physiology of mother in, 414
Premotor cortex, 259, 262
Pressoreceptor system, 159
Pressure
  at sea depths, 208
  relationship to flow, 134
Pressure gradient across capillary membrane, 53
Pressure pulses
  damping of, 155
  transmission to small vessels, 155
Pressurization at high altitudes, 203
Proconvertin, 111
Progesterone, 404
  in pregnancy, 411
Progressive shock, 173
Proprioceptor reflexes, 272
Proprioceptors, 248
  equilibrium and, 278
Proteins, 334
  absorption of, 342
  conversion of
    to amino acids, 345
    to carbohydrates, 346
  digestion of, 342
  hydrocortisone and, 381
  in ATP synthesis, 349
  in fat synthesis, 340
  in lymph, 57
  in protoplasm, 14
  in tissues, 344
  insulin and, 384
  partial, 351
  pinocytosis of, 24
  plasma, formation of, 344
  requirements in body, 351
Protein synthesis, 27, 344
  control of, 33, 34
  mechanism of, 28
  summary of, 34
Prothrombin,decreased, bleeding and, 111
Prothrombin time, 111
Protoplasm
  carbohydrates in, 15
  chemical nature of, 14
  colloidal nature of, 20
  nucleoproteins in, 15
Psychosomatic diseases, 292
Psychosomatic effects
  transmission of
    through autonomic nervous system, 292
    through glands, 292
    through muscles, 292
Psychosomatic pain, 293
Psychosomatic physiology, 280, 291
Ptyalin, 335
Puberty, 400
Pulmonary capillaries, dynamics of, 143
Pulmonary circulation, 141
  pressures in, 142
Pulmonary congestion, 143
  in heart failure, 171

Pulmonary edema, 198
in heart disease, 127
Pulmonary membrane, 185
Pulmonary vessels, as blood reservoir, 140
Pulp of teeth, 393
Pulse pressure
abnormal, 154
factors affecting, 154
Pupil of eye, control of, 308
Purkinje system, 121
Putamen, 265
Pyridoxine, 354

Radiation, for heat loss, 362
Radiation hazards in spaceships, 208
Reciprocal inhibition, 274
Rectum, reflexes controlling, 275
Red blood cells, 87
formation of, 89
fragility of, 92
genesis of, 89
life of, 91
organization of, 88
regulation of production, 91
Referred pain, 252
Reflex(es)
autonomic, of spinal cord, 275
axon, 271
crossed extensor, 274
duodenocolic, 327
extensor thrust, 273
flexor, 273
gastrocolic, 327
heat, 285
Hering-Breuer, 192
micturition, 74
mucosal, 327
muscle, 272
nociceptive, 273
of spinal cord, 271, 274
pain, 273
peritoneal, 275, 327
proprioceptor, 272
scratch, 274
stretch, 272
taste, 316
tendon, 273
walking, 275
withdrawal, 273
Reflex arc, 236
Refractory period
of heart, 122
of nerve fibers, 218
Relaxing factor, 227
Renal. See also *Kidney.*
Renal clearance, 70
Renal hypertension, 162
Renal tubules
absorption from, 67
active reabsorption in, 68
active secretion in, 69
collecting, urine concentration and, 77
diffusion in, 68
Renin, 160
Repolarization of nerve membrane, 217
Reproduction, 407
of cells, 32, 37
Reproductive system, overall function in body, 10
Residual volume, 182
Resonance of sound in cochlea, 311
Respiration, 177
artificial, 180
cellular, 25
during speech, 195
mechanics of, 177
multiple factor theory for, 195
muscles of, 180
regulation of, 191
by carbon dioxide, 193
by hydrogen ion concentration, 193
by oxygen deficiency, 194
rhythm of, basic, 191
spirometry and, 182
Respirator, artificial, 181
Respiratory abnormalities, 191, 195
Respiratory center, 191
failure of, 192
Respiratory rate, 182
Respiratory system
integration in body, 7
regulation of acid-base balance, 79
Respiratory volume, minute, 182
Respirometer, 360
Resuscitator, 181
Reticular activating system, 287
stimulation by cortex, 288
Reticuloendothelial system, 100
Reticulum cells, 100
Retina, 300
connections with brain, 303
Reverberating circuits, 243
Reward and punishment, learning and, 291
Rh factor, 112
Rhodopsin, 301
Rhythmicity
in heart muscle, 120, 233
in smooth muscle, 234
Riboflavin, 354
Ribonucleic acids, 27
control of activity, 35
in protein synthesis, 27
production by genes, 33
repression of, 35
Ribosomes, 17
Rickets, 389
Rickettsia, 13
Rods, of retina, 300

S-A node, as pacemaker of heart, 120
Saline, as blood substitute, 114
Salivary secretion, 328
Sarcolemma, 223
Scratch reflex, 274
SCUBA diving, 210
Secretion
by liver, 331
by pancreas, 330
by small intestine, 332
in gastrointestinal tract, 321
of digestive juices, 282
salivary, 328
Secretory granules, 18
Segmentation in small intestine, 325
Semicircular canals, 278
Sensations
affective nature of, 249
discrimination of intensity of, 251
interpretation by brain, 254
localization on body, 249

Sensations (*Continued*)
modality of, 249
transmission to brain, 246
Sensory cortex, 250
Sensory receptors, 248
Sex characteristics, male, testosterone and, 401
Sex determination by sperm, 398
Sex hormones
female, 403
male, 400
Sexual act
female, 403
male, 399
Sexual climax
in female, 403
in male, 399
Sexual cycle, female, regulation of, 404
Sexual functions, 397
autonomic nerves and, 284
female, 397, 401
male, 397
Shivering, 365
Shock, circulatory, 164, 171
stages of, 172
treatment of, 174
Skeletal muscle, 225
anatomy of, 225
tone of, 232
Skeletal muscle contraction, 226
Skeleton
integration in body, 4
lever systems of, 230
positive acceleration and, 205
Skin
blood flow through, 150
nutrient, 151
Sleep, 280, 287
effects of, 289
feedback theory of, 288
Small intestine. See *Intestine, small.*
Smell, 309, 316
and taste sensation, 316
primary cortex for, 255
primary sensations of, 317
transmission of sensations into brain, 318
Smooth muscle. See *Muscle, smooth.*
Sneeze, 179
Sodium
active transport of, 24
aldosterone and, 380
regulation in body fluids, 75
Sodium pump, 24
Solutes, diffusion to cells, 46
Somesthetic association area, 254
Somesthetic cortex, 250
Somesthetic sensations, 246
Sound
characteristics of, 309
conversion to nerve impulses, 313
determination of loudness, 312
determination of pitch by cochlea, 312
resonance in cochlea, 311
Space physiology, 200, 206
Spatial summation, 222
Speech
control of, 268
respiration in, 195
Speech center, 269
Sperm, 397
entry into ovum, 407
storage of, 399
Spermatogenesis, 398
testosterone and, 400
Spinal anesthesia, arterial pressure and, 158
Spinal cord
control of blood vessels, 275
physiologic anatomy of, 270
reflexes of, 271
in body temperature regulation, 366
rhythmic, 274
transmission of motor signals, 263
Spirogram, 182
Spirometry, 182
Spleen
as blood reservoir, 139
phagocytic function of, 149
reservoir function of, 149
Stagnant anoxia, 195
Starches, digestion of, 335
Starvation, 357
Sterility
female, 403
male, 400
Stimuli for nerve fibers, types of, 218
Stomach
mixing in, 325
motor functions of, 324
regulation of emptying, 325
secretion in, 329
neurogenic control of, 329
phases of, 330
storage by, 324
Stretch reflex, 272
damping function of, 273
Stroke(s), 148
Stroke volume output, 154
Subthalamic nucleus, 266
Sugar, digestion of, 335
Summation
at synapse, 238
in nerve fibers, 222
of muscle contraction, 231
Surface active agent, 180
Surface precipitation reaction, 16
Swallowing, 323
Sweat glands, sympathetic stimulation and, 282
Sweating, 362, 365
Sympathetic nerves. See *Nerve(s), sympathetic.*
Sympathetic nervous system. See *Nervous system, sympathetic.*
Sympathetic stimulation, blood glucose and, 337
Sympathetic tone, 285
Synapse, 236
anatomy of, 237
conduction at, one-way, 240
facilitation at, 239
fatigue of, 240
function of, 237
memory and, 240
transmission of impulses at, 238
Synthesis in cells, ATP in, 26
Systemic circulation
blood flow in, 131
pressures at different points, 156
Systemic pressure, mean, venous return and, 166
Systole, 127
Systolic pressure, 152

Taste, 309, 315
diet regulation by, 316
primary cortex for, 255
primary sensations of, 315

Taste bud, 315
Taste reflexes, 316
Teeth, 393
  formation of, 394
  metabolism of, 394
  mineral exchange in, 395
  physiology of, 387, 393
Temperature of body, 358, 363
  cold, thyroxine and, 377
  effect on body, 367
  normal values, 363
  skin blood flow and, 150
Temporal summation, 222
Tendon reflex, 273
Testosterone, 400
Tetanization, 231
Tetany, 388
Tetralogy of Fallot, 144
Thalamus
  consciousness and, 249
  cortex and, 256
  emotions and, 291
Thiamine, 352
Thirst, mechanism of, 83
Thought processes, 257
  nature of, 257
  patterns in brain, 259
Thrombin, 108
Thrombocytopenia, 110
Thromboplastin, 108
Thymus, transplants and, 115
Thyroid hormone
  fat and, 341
  in regulating body temperature, 365
  metabolism and, 359
Thyrotropin, 373
Thyroxine, 371, 375
  effects on body, 376
  mechanism of action, 376
  regulation of production, 376
Tidal volume, 182
Tissue(s)
  fat, 339
  fluid flow in, 55
Tissue histiocytes, conversion from lymphocytes, 98
Tissue immunity, 103
Tissue pressure, 53
  negative value of, 58
  positive in edema, 59
  regulation by lymphatics, 57
Tissue transplant, autologous, 115
Tone
  in skeletal muscle, 232
  in smooth muscle, 234
  sympathetic, 285
Toxoid, 104
Transfusion, 107, 111
Transfusion reactions, kidneys and, 113
Transplantation of organs, 107, 115
Trophoblasts, in fetal nutrition, 408
Tympanic membrane, 310

Unesterified fatty acids, 339
Unmyelinated fibers, 221
Urea, lack of reabsorption in tubules, 69
Ureters, 73
Urinary bladder, 73
  autonomic nerves and, 284
  emptying of, 74
  spinal cord control of, 275
Urinary system, 41
Urine
  formation of, 64
  storage in bladder, 73
Urticaria, 105
Uterus, implantation of ovum in, 408
Utilization coefficient, 188
Utricle, 277

Vaccination, 104
Vascular obstruction, as cause of shock, 172
Vascular tone, 138
Vasomotor center, 158
Vasomotor tone, 157
  cardiac output and, 167
Vasopressin, 375
Veins
  as blood reservoir, 139
  circulation in, 146
  in cerebral circulation, 147
Venous pooling in shock, 172
Venous pressure, 164, 169
  in pulmonary circulation, 142
  measurement of, 170
Venous pump, 169
Venous return, 165
  diminished in shock, 172
  equilibrium with cardiac output, 167
Venous sinuses in cerebral circulation, 147
Ventricular fibrillation, 124
Vestibular apparatus, 277
Viral physiology, 3
Virus(es), 13
Visceral pain
  qualities of, 253
  stimulus for, 252
Visceral sensations, 251
Visceral sensory fibers, 281
Vision
  chemistry of, 301
  color, 303
  fields of, 304
  fusion of, 306
Visual acuity, 305
Visual cortex, 254, 304
Vital capacity, 182
Vitamin A, 352
Vitamin B_{12}, 354
  in formation of red cells, 90
Vitamin D, 355
  in calcium metabolism, 388
  source of, 388
Vitamin E, 355
Vitamin K, 355
  deficiency of, bleeding in, 111
Vitamins
  in foods, 353
  requirements of body, 352
Vitreous humor, 297
Vocal cords, function of, 178
Voice, 178
Vomiting, 327

Wakefulness, 287
  effects of, 289
  feedback theory of, 288
Walking reflexes, 275

Water
  absorption of, 346
  as cause of diuresis, 78
  diffusion of
    through membranes, 22
    to cells, 46
  in body, 43
  in circulation, 50
  in protoplasm, 14
  thirst mechanism for control of, 83
Water vapor, effects at high altitude, 202
Wave summation, in muscle contraction, 231
Weber-Fechner law, 251
Weightlessness, 206
White blood cells, 94
  differential count of, 98
  genesis of, 94
  life of, 94
Withdrawal reflex, 273

Zinc, 356

Middlebury College
0 00 02 0808819 5